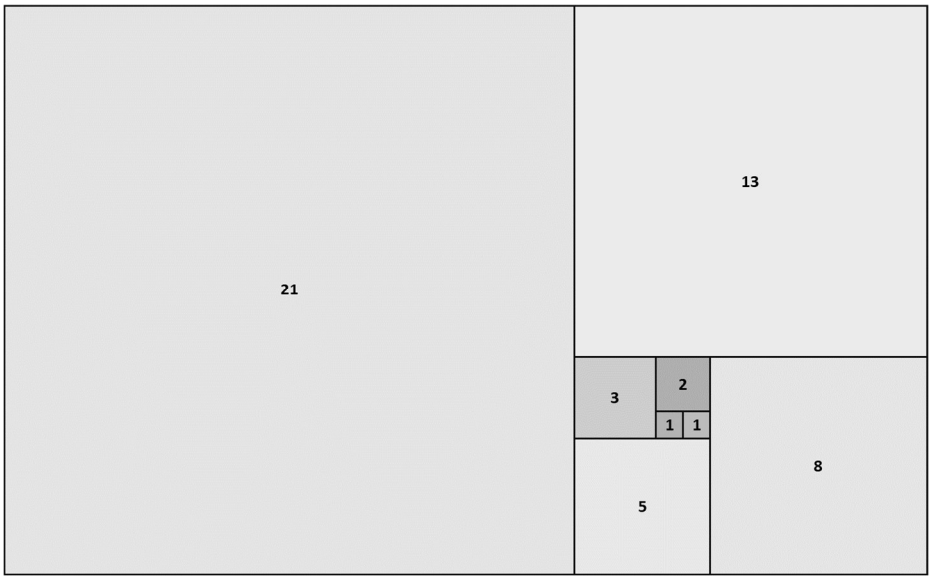
Mathematical Thinking

2nd Edition



1,1,2,3,5,8,13,21,?

$$F_n = F_{n-1} + F_{n-2}$$

by **Stephen Fratini**

Table of Contents

List of Figures

List of Tables

Preface

What is mathematics and why should college students in areas other than Science, Technology, Engineering and Mathematics (STEM) be required to take mathematics courses in college? As a graduate student and then briefly as a visiting assistant professor, I taught math at the college-level to both STEM and non-STEM students. I often got the question (especially from non-STEM majors), "why am I taking this class, and of what use will it be when I get a job in industry?" At the time, my best answer was that the class will help you think better, improve your problem-solving skills and some of the material will be directly applicable (e.g., continuous interest). My stint as a mathematics teacher (some 7 years in total) was followed by 31 years as an engineer in the telecommunications industry, and to be honest, rarely did I directly use my mathematical skills (key word here is "directly"). The same was true for most of my colleagues (many of whom were engineers with significant backgrounds in applied mathematics). However, what I did use (and valued quite a lot) were the problem solving, modeling and analytical skills that I developed during my time as a mathematician. This all led me to think about authoring a book to help students (advanced high-school or college) and other interested folks to improve their basic analytical skills while at the same time getting a small taste of many different aspects of mathematics.

I asked several colleagues (all non-academics) for their opinions on what topics from mathematics would be useful for such a book. I even posted a question on my LinkedIn feed. The responses centered around probability and statistics, financial math and computer programming. The suggestion about computer programming sent me into another line of thinking. While I didn't want to author a book on computer programming, I did see value in having some coverage of algorithmic thinking (see Section 17). This thought, in turn, led me to consider drafting a book on "mathematical thinking," with algorithmic thinking being one of the topics. The idea was to cover the thought processes, structures and models behind mathematics, e.g., logic, methods of proof and methods of reasoning, and in fact, "mathematical thinking" is at the core of this book.

The proofs of theorems are included to develop the reader's thought processes. So, please do not skip the proofs. On the other hand, and to be honest, there are a few places where I went "off the deep end" in terms of details, i.e., the axiomatic definition of sets and the proof of the associative law for Boolean algebras.

An outline of the book is provided in the introduction, along with a statement of dependencies among the sections. The sections of the book are intended to be read in order.

In terms of writing style, the book is written in a combination of the passive voice, with some use of the first person plural (as is typical in mathematical papers). In addition to this preface, there are a few other places where I state personal opinions. Such statements are prefaced by the phrase "Author's Remark." This reminds me that I once sent a paper to a journal for review with a similar mixed writing style and one of the reviewers went on a tirade. I really don't see the issue. In the words of Ralph Waldo Emerson:

> A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines. With consistency a great soul has simply nothing to do. He may as well concern himself with his shadow on the wall. Speak what you think now in hard words, and tomorrow speak what tomorrow thinks in hard words again, though it contradicts everything you said today. — "Ah, so you shall be sure to be misunderstood." — Is it so bad, then, to be misunderstood? Pythagoras was misunderstood, and Socrates, and Jesus, and

> Luther, and Copernicus, and Galileo, and Newton, and every pure and wise spirit that ever took flesh. To be great is to be misunderstood.

Second Edition: The main reason for the second edition was to make all the figures black and white so as to reduce the publication cost on Amazon. There are also some minor corrections.

Acknowledgments

I'd like to thank the following people for their assistance in reviewing a draft of this book: Laura Bagwell, Michael Brenner, Norman Dorn, Zach Gilstein (special thanks for reviewing the entire book), Vaidyanathan Ramaswami, and Wendy Teller. Their efforts led me to make several additions and improvements.

Stephen Fratini
Sole Proprietor of The Art of Managing Things
Eatontown, New Jersey (USA)
Email: sfratini@artofmanagingthings.com
LinkedIn: www.linkedin.com/in/stephenfratini

1 Introduction

1.1 Purpose

The purpose of this book is to improve the reader's analytical skills through the study and practice of "mathematical thinking" where "mathematical thinking" includes algorithms, logic, methods of reasoning, methods of proof, modeling, and universal mathematical laws. As a byproduct, the reader is provided with a brief introduction to many areas of mathematics.

1.2 Intended Audience

The intended audience includes students (advanced high school and college) and folks in general who are interested in improving their analytical thinking skills and at the same time learning some mathematics. For those who don't deal with mathematics on a regular basis, this will not be an easy read but hopefully, the benefits will be worth it.

1.3 Prerequisites

The prerequisites are fairly basic, i.e., high school algebra, a little bit of basic geometry, and some prior exposure to mathematical proofs. Most of the topics in the book are developed from basic principles.

1.4 Terminology

The document has a large amount of notational shorthand, most of which is introduced when first used. There are, however, a few overall notations that are best introduced up front in this introduction, i.e.,

- In terms of notation, the proofs are ended with the symbol ■ which should be interpreted as "which was to be proved."
- Mathematicians often use the statement "if and only if" as a shorthand. For example, "Statement A is true if and only if Statement B is true." This is a shorthand for the equivalent statements: "If Statement A is true, then Statement B is true. If Statement B is true, then Statement A is true."
- When quoting a longer selection of text, a block quotation is used. A block quotation is a direct quotation that is not placed inside quotation marks but instead is set off from the rest of the text by starting it on a new line and indenting it from the left margin.

1.5 Outline

Table 1 provides an outline of the book, with a list of dependencies for each section. In some cases, the dependencies need to be traced back several steps. For example, the section on statistics depends on the section on probability which in turn, depends on the sections covering combinatorics and calculus, and so on. In general, the sections are designed to be read in order.

Table 1. Outline of Book

Section Title	Contents	Dependencies
Introduction	Purpose, intended audience, prerequisites	none
What is Mathematics?	Short section concerning definition of mathematics	none
Methods of Reasoning	Short section concerning general methods of reasoning	none
Propositional Logic	Basic logic notation and concepts, conversion of arguments to logic statements and proof of validity	none
First-Order (or Predicate) Logic	Builds upon propositional logic, logic paradoxes	Propositional logic
Sets	Basic theory of sets	none
Boolean Algebra	A Boolean algebra is a general structure of which logic and sets are examples	Helpful to have read the sections on logic and sets
Functions	Basic concepts of functions including graphs, polynomials, exponential functions, logarithmic functions, and transformations of functions	none
Number Theory	Basic concepts of divisibility, mathematical induction, binomial theorem, prime numbers, factorization of integers	none
Combinatorics	Techniques and concepts concerning counting	Sets, number theory
Calculus	Very brief introduction to differential and integral calculus	Functions
Probability	Various ways of stating odds, computation of probabilities, conditional probability, Bayes' theorem, random variables, law of large numbers, central limit theorem	Combinatorics, calculus
Statistics	Descriptive statistics, inferential statistics, hypothesis testing, statistical paradoxes	Probability
Graph Theory	Basic concepts, classification of graphs, connectivity, minimum spanning trees, traversing graphs	Mathematical induction from the number theory section
Linear Algebra	Matrices, vectors, systems of equations, vector spaces, linear independence, linear transformations, bases	Functions
Proofs	Summary of the different types of proofs, drawing on examples from earlier in the book	Other sections in the book that involve proofs
Algorithms	Basic concepts, classification of algorithms, graphical representation (e.g., flow charts)	Makes use of some algorithms used earlier in the book
Universal Laws of Mathematics	Law of large numbers, central limit theorem, Benford's law, power laws	Probability, statistics

Section Title	Contents	Dependencies
Conclusion	Brief wrap-up of the ideas presented in the book and some suggestions for further exploration	
Acronyms and Symbols		
References		
Index of Terms		

2 What is Mathematics?

Most dictionaries define mathematics by listing some of the constituent parts, e.g., The Free Dictionary by Farlex provides the following definitions:

> (Mathematics) (functioning as singular) a group of related sciences, including algebra, geometry, and calculus, concerned with the study of number, quantity, shape, and space and their interrelationships by using a specialized notation.
>
> (Mathematics) (functioning as singular or plural) mathematical operations and processes involved in the solution of a problem or study of some scientific field.

From Wikipedia article on mathematics:

> Mathematics (from Greek μάθημα máthēma, "knowledge, study, learning") includes the study of such topics as quantity (number theory), structure (algebra), space (geometry), and change (mathematical analysis). It has no generally accepted definition.

The above definitions are not particularly satisfying or helpful beyond giving one an idea of the topics that come under mathematics.

The Encyclopaedia Britannica provides the following definition [1]:

> Mathematics, the science of structure, order, and relation that has evolved from elemental practices of counting, measuring, and describing the shapes of objects.

[**Author's Remark:** Perhaps "less is more," regarding the above definition. I would have preferred just "Mathematics is the science of structure, order, and relation." One of my mathematics teachers from college, who specialized in mathematical logic, defined mathematics as "the study of logical structure" which fits in well with the definition from Encyclopedia Britannica.

In my view, an attorney engaging in precise cross-examination of a witness is using mathematical thinking. Similarly, a poet writing a structured poem is thinking mathematically. My point is that non-mathematicians and non-scientists do require structured thought processes that would benefit from studying mathematics.]

3 Methods of Reasoning

3.1 Deductive Reasoning

From the Free Dictionary by Farlex:

> Deductive reasoning – reasoning from the general to the particular (or from cause to effect)
>
> - abstract thought, logical thinking, reasoning – thinking that is coherent and logical
> - syllogism – deductive reasoning in which a conclusion is derived from two premises.

Deductive reasoning entails the extraction of conclusions from multiple premises where a logical relationship is constructed between the propositions and the conclusion. When the premises are true and the rules of deduction (basic logic) are properly applied, the resulting conclusions are undeniable true.

In the context of this book and for mathematics in general, deductive reasoning entails the derivation of truths from a collection of definitions and assumed rules or principles (known as axioms).

In Euclid's Elements [2] (a textbook on geometry dating back to about 300 BC), the author deduces 465 theorems and geometric constructions from only 5 axioms, 5 common notions and 131 definitions.

This book is mainly focused on deductive reasoning, with the exception of the section on Statistics which provides processes ("tools") to assist with inductive reasoning.

[Author's Remark: To be clear, mathematics includes both deductive and inductive reasoning. It is only a matter of preference that I have chosen to focus more on the deductive aspects of mathematics in this book.]

3.2 Inductive Reasoning

From the Free Dictionary by Farlex:

> Inductive reasoning – reasoning from detailed facts to general principles
>
> - generalization, induction
> - colligation – the connection of isolated facts by a general hypothesis.

Inductive reasoning entails a process in which specific instances or situations are observed and analyzed with the intent of determining general principles (e.g., physical laws) or trends (e.g., who may win an election). In this process, the multiple observations or experiments are believed to provide convincing evidence for the truth of the conclusion. Inductive reasoning is used to develop an understanding, on the basis of observing regularities, to ascertain some knowledge about the world. However, there is a possibility that the conclusions may be false, in contradiction to what has been witnessed in a relatively limited set of observations and experiments.

As noted, this book is mainly focused on deductive reasoning. That is not to say that the process of doing mathematics is purely deductive. For example, the process of determining an appropriate set of definitions and axioms for a given aspect of mathematics is more of an exploratory process that is more inductive than deductive.

The discovery of physical laws is an inductive process. The physicist observes many events or performs experiments with the goal of getting insights into natural phenomena. The result of such a process may lead to the determination of a physical law, e.g., $F = ma$ (force equal mass times acceleration). From a collection of physical laws, the physicist may derive other truths via deductive reasoning. So, inductive and deductive reasoning can be used together in physics as well as in other sciences.

In other cases, inductive reasoning may not lead to a principle or law. For example, a pollster may take several polls to help predict the outcome of an election. The conclusion of the inductive reasoning related to the polling is a prediction.

3.3 Comparison

Table 2 provides a comparison between deductive and inductive reasoning.

Table 2. Comparison of Deductive and Inductive Reasoning

	Deductive Reasoning	**Inductive Reasoning**
Summary	Deductive reasoning (or top-down logic) entails a thought process (guided by principles of logic) that leads from general statements regarding what is known or assumed true (axioms) to conclusions which are necessarily true.	Inductive reasoning (bottom-up logic) entails a thought process (often guided by the principles of statistics) that constructs or evaluates general propositions that are derived from specific examples.
Arguments	Arguments in deductive logic are either valid or invalid. An argument is valid if and only if it takes a form that makes it impossible for the premises to be true and the conclusion nevertheless to be false. (A formal definition of a valid argument is provided in Section 4.9.1.)	Arguments in inductive reasoning are either strong or weak. The strength or weakness of arguments depends on the quality and quantity of observations and experiments, and the competency in which the observations and experiments are analyzed.
Validity of conclusions	Conclusions can be proven to be valid if the premises are known to be true.	Conclusions may be incorrect even if the arguments are strong and the premises are true.
Process	Existing Definitions, Axioms and Proven Theories → Hypothesis → Logical Deduction → Proven Hypothesis (i.e., a new theorem)	Experimentation and Observation → Identify Patterns and Trends → Tentative Hypothesis → Statistically Supported Theory or Trend (there is the possibility of looping back from the tentative hypothesis to experimentation and observation)
Structure	Goes from general to specific	Goes from specific to general
Draws inferences with	Logic	Statistical analysis

4 Propositional Logic

4.1 Overview

4.1.1 Definition

This section covers something called propositional logic, and is perhaps the most basic of all the material in this document. According to the Internet Encyclopedia of Philosophy, propositional logic is defined as follows [3]:

> Propositional logic, also known as sentential logic and statement logic, is the branch of logic that studies ways of joining and/or modifying entire propositions, statements or sentences to form more complicated propositions, statements or sentences, as well as the logical relationships and properties that are derived from these methods of combining or altering statements. In propositional logic, the simplest statements are considered as indivisible units, and hence, propositional logic does not study those logical properties and relations that depend upon parts of statements that are not themselves statements on their own, such as the subject and predicate of a statement ...

"The box is heavy" and "The box is green" are example propositions. If the letter A is used to represent the first statement and B to represent the second statement, then we can join A and B in various ways, e.g., A and B ("The box is heavy" and "The box is green"). This section covers various logical connectives and their associated properties.

4.1.2 Motivation

Studying logic is like learning a new language, albeit one with a small vocabulary and just a few rules of grammar. The language of logic flows through almost all of mathematics. Further, logical reasoning is required in all fields of endeavor (going well beyond mathematics and the sciences).

For example, consider the following argument:

> In order to be good at basketball, one needs to practice shooting, do resistance and aerobic training, and study the various defensive and offensive strategies. Sally practices shooting, and does engage in resistance and aerobic training but does not know well the defensive and offensive strategies. Thus, Sally is not a good basketball player at this point.

How would one prove the above argument is valid? In this section, the framework to logically determine the validity of this and other arguments will be developed.

Some other advantages of learning logic include:

- Logic can be used as an effective tool of persuasion.
- An understanding of logic helps you to identify fallacies.
- Logic is fun. If you like solving puzzles, you should also like solving logic problems.

4.2 Basic Logical Operations and Definitions

In the realm of propositional logic, a **proposition** (or statement) is a declarative sentence or phrase which is either true or false. In what follows, a proposition is represented by a letter, e.g.,

- A: Practice shooting a basketball

- B: Do resistance and aerobic training
- C: Study the various defensive and offensive strategies of basketball
- D: Be good at basketball

It is possible to combine propositions to form compound propositions. A proposition that is not composed of simpler propositions is called a primitive proposition (or just a primitive).

Propositions can be combined using the following operations:

- **Conjunction (AND)**: For two propositions (label them as A and B), their conjunction is represented as $A \wedge B$. The proposition $A \wedge B$ is defined to be true if and only if proposition A and proposition B are true.
- **Disjunction (OR)**: For two propositions (label them as A and B), their disjunction is represented as $A \vee B$. The proposition $A \vee B$ is defined to be true if and only if either
    - proposition A or proposition B is true, or
    - both proposition A and proposition B are true.

  (This is different from "exclusive OR" which is false if both A and B are true.)
- **Negation**: The negation of proposition A is represented as $\neg A$. The proposition $\neg A$ is defined to be true if A is false, and the proposition $\neg A$ is defined to be false if A is true.

For example, the sentence "In order to be good at basketball, one needs to practice shooting, do resistance and aerobic training, and study the various defensive and offensive strategies" can be written more formally as "$A \wedge B \wedge C$ implies D" using the assignments for A, B, C and D above.

As another example, consider the following propositions:

- D: $2 + 2 = 5$. (false)
- E: New York city is in Alaska. (false)
- F: $3 \times 9 = 27$. (true)

Given the above assignments, $D \wedge E$ is false, $D \wedge F$ is false, $D \vee F$ is true and $\neg D \wedge \neg E$ is true.

The following two terms are commonly used in propositional logic:

- **Tautologies** are propositions that are always true, e.g., $A \vee \neg A$ is always true.
- **Contradictions** are propositions that are always false, e.g., $A \wedge \neg A$ is always false.

4.3 Truth Tables

When studying the composition of propositions, a common problem is to determine when a compound proposition is true or false based on the constituent propositions. For example, when is the proposition $(A \wedge B) \vee C$ true or false for given values of A, B and C? One approach is to use something called a **truth table**. As an example, consider the truth table in Table 3. On the left, we list all possible combinations of values for propositions A and B. On the right, we list the values for $A \vee B$ and $A \wedge B$. For example, the third row of the table tells us that if A is false (F) and B is true (T), then $A \vee B$ is true and $A \wedge B$ is false. Table 3 can be used as an alternate way of defining conjunction and disjunction.

Table 3. Truth Table for Conjunction and Disjunction

A	B	$A \vee B$	$A \wedge B$
T	T	T	T
T	F	T	F
F	T	T	F
F	F	F	F

Returning to the question about when the proposition $(A \wedge B) \vee C$ is true or false, it is usually best to divide more complex propositions into parts, as is shown in Table 4. First, $A \wedge B$ is computed (noting that it does not depend on the value of C) and then $(A \wedge B) \vee C$ is computed by doing an "or" operation on columns 4 and 3 in the table. Also, note that there are 8 possible combinations of T and F for the three propositions A, B and C.

Table 4. Truth Table for $(A \wedge B) \vee C$

A	B	C	$A \wedge B$	$(A \wedge B) \vee C$
T	T	T	T	T
T	F	T	F	T
F	T	T	F	T
F	F	T	F	T
T	T	F	T	T
T	F	F	F	F
F	T	F	F	F
F	F	F	F	F

Various combinations of propositions can give rise to the same truth table. In such cases, the propositions are said to be **equivalent**. If A and B are equivalent, we write $A = B$. Table 5 shows the truth table for equivalent (and for "not equivalent"). In words, the proposition $A = B$ is true if and only if A and B are either both true, or both false. The proposition $A \neq B$ is true if and only if A and B have different values.

It should also be clear from Table 5 that $(A = B) = \neg(A \neq B)$.

Table 5. Truth Table for Equivalent and Not Equivalent

A	B	$A = B$	$A \neq B$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	T	F

4.4 Derived Operations

From the three basic operations defined in Section 4.2, it is possible to derive other operations. Some of the more common examples are as follows:

- Exclusive OR (XOR) – A or B but not both. This can be represented as $A \ \boldsymbol{XOR}\ B = (A \vee B) \wedge \neg(A \wedge B)$.
- "NOT AND" (NAND) – the negation of A and B, i.e., $A \ \boldsymbol{NAND}\ B = \neg(A \wedge B)$. NAND is sometimes referred to as the Sheffer stroke or alternative denial. The Sheffer stroke is written as the vertical stroke (i.e., $A \mid B$) or up pointing arrow (i.e., $A \uparrow B$).
- "NOT OR" (NOR) – the negation of A or B, i.e., $A \ \boldsymbol{NOR}\ B = \neg(A \vee B)$. NOR is sometimes referred to as the joint denial operation and written with a down pointing arrow, i.e., $A \downarrow B$.

4.5 Conditional and Biconditional Propositions

One proposition can imply another. For example, take the propositions

- A: "The sun is shining"
- B: "It is bright outside."

Proposition A implies Proposition B. The notation for "implies" is $A \Rightarrow B$. This is also phrased as "If A, then B."

$B \Rightarrow A$ is called the **converse** of $A \Rightarrow B$. It is possible for $A \Rightarrow B$ to be true and the converse to be false.

$\neg B \Rightarrow \neg A$ is called the **contrapositive** of $A \Rightarrow B$.

If the implication is bidirectional (i.e., A implies B, and B implies A), then we write $A \Leftrightarrow B$. The proposition $A \Leftrightarrow B$ can also be phrased as "A if and only if B" or sometimes a shorter version is used, i.e., "A iff B". Clearly, $A \Leftrightarrow B$ is equivalent to $(A \Rightarrow B) \wedge (B \Rightarrow A)$.

The truth table for the conditional and biconditional operations is shown in Table 6.

Table 6. Truth Table for Conditional and Biconditional

A	B	$A \Rightarrow B$	$A \Leftrightarrow B$
T	T	T	T
T	F	F	F
F	T	T	F
F	F	T	T

As can be seen from Table 7, $A \Rightarrow B$ and it contrapositive $\neg B \Rightarrow \neg A$ are equivalent.

Table 7. Truth Table for Equivalent Conditional Propositions

A	B	$\neg A$	$\neg B$	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

So, if the proposition A is false, the statement $A \Rightarrow B$ is true (regardless of whether B is true or false). One may ask "why is that the case?" To make this issue more concrete, consider the following example.

- A: Frogs can understand logic.
- B: $\frac{14}{2} = 7$.
- C: $2 + 2 = 7$

$A \Rightarrow B$ is true. In words: "**If** 'Frogs can understand logic' then the $\frac{14}{2} = 7$" is a true statement.

Also, $A \Rightarrow C$ is true. In words: "**If** 'Frogs can understand logic' then the $2 + 2 = 7$" is a true statement.

One explanation is to say the $A \Rightarrow B$ is the same as (i.e., has the same truth table as) $(\neg A) \vee B$. While this is true, it also begs the question "why is $A \Rightarrow B$ defined to have the same truth table as $(\neg A) \vee B$?"

Another approach is to focus on the "if" part of the statement $A \Rightarrow B$. In the examples above, the "if" is in bold to emphasize that the first part of those two conditional statements is not going to happen. Both statements (i.e., $A \Rightarrow B$ and $A \Rightarrow C$) are true, because neither statement promised anything that will ever happen (i.e., A will never happen). One can say that the statements are vacuously true. [The concept of a statement being "vacuously true" is used in logic and set theory. The concept applies when a statement attributes a property to all members of an empty set. For example, "all pigeons on Mars are green" is vacuously true.]

4.6 Redundancy

In the previous section, we saw that $A \Rightarrow B$ has the same truth table as $(\neg A) \vee B$ and so, one could argue that "implies" is redundant since it can be represented using negation and disjunction. We can even obviate the need for disjunction by using conjunction and negation, i.e., $A \vee B = \neg(\neg A \wedge \neg B)$. Is there a single logic operation that can be used to represent all binary logic operations? The question is important from both a theoretical point of view and from a practical point of view (since implementations of logic functions are used in computers).

To answer the question, we first note there are 16 possible outcome arrangements for binary logic operations. This can be seen from Table 8, the arbitrary operation (o) has two possible values for each pair of values for A and B. Thus, we get 2 x 2 x 2 x2 = 16 possible truth tables (and associated operations).

Table 8. Truth Table for Binary Operations

A	B	$A\ o\ B$
T	T	2 possible values (T or F)
T	F	2 possible values (T or F)
F	T	2 possible values (T or F)
F	F	2 possible values (T or F)

In fact, all 16 of the possible operations can be generated by combinations of just NANDs or just NORs. This fact was proven by Henry Sheffer in 1913 [4]. However, some of the equivalences are not simple. For example, the following arrangement of NANDs is needed to generate the same truth table as NOR, assuming A and B are the input propositions:

$$[\,(A \uparrow A) \uparrow (B \uparrow B)] \uparrow [\,(A \uparrow A) \uparrow (B \uparrow B)].$$

It is also possible to generate all 16 possible operations using various combinations of just two operations, e.g., conjunction and negation. The section on Redundancy, in the Wikipedia article entitled "Logical connective" [5], lists all the various combinations of logical operations that can be used to generate the entire set of 16 binary logic operations.

From a practical point of view, NAND and NOR logic gates are physical implementations of NAND and NOR operations in Integrated Circuits (ICs). NAND and NOR logic gates are used extensively in IC implementation because they are easy and economical to fabricate.

4.7 Logical Laws

Various laws hold true for combinations of the logical operations covered in the previous subsections. The following theorem summarizes some of the most common and useful laws for logical operators.

Theorem 4-1 The following laws hold true

- $A \vee \neg A = T$ (example of a tautology)
- $A \wedge \neg A = F$ (example of a contradiction)
- $A \vee A = A$ and $A \wedge A = A$ (Idempotency)
- $A \vee B = B \vee A$ and $A \wedge B = B \wedge A$ (Commutativity)
- $(A \vee B) \vee C = A \vee (B \vee C)$ and $(A \wedge B) \wedge C = A \wedge (B \wedge C)$ (Associativity)
- $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$ and $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$ (Distributive laws)
- $\neg(A \vee B) = (\neg A) \wedge (\neg B)$ and $\neg(A \wedge B) = (\neg A) \vee (\neg B)$ (De Morgan's Laws)

Proof: All of the above can be proven by showing that both sides of a given equation have the same truth table. Proof of the first De Morgan's laws is shown below.

Table 9. Truth Table concerning De Morgan's Laws

A	B	$\neg A$	$\neg B$	$\neg \boldsymbol{A} \wedge \neg \boldsymbol{B}$	$A \vee B$	$\neg(\boldsymbol{A} \vee \boldsymbol{B})$
T	T	F	F	**F**	T	**F**
T	F	F	T	**F**	T	**F**
F	T	T	F	**F**	T	**F**
F	F	T	T	**T**	F	**T**

Equality of the two columns in bold font prove the desired result■

4.8 Determining a Statement that Satisfies a Truth Table

In some cases, one may have knowledge of a particular truth table but not the original statement that corresponds to the truth table. For example, consider a logical circuit ("black box" as shown in Figure 1) where we know all outputs for each combination of inputs but don't know the logical expression (internals of the "black box") to which the circuit corresponds. There is a technique for working backwards but it only gives one of many possible solutions.

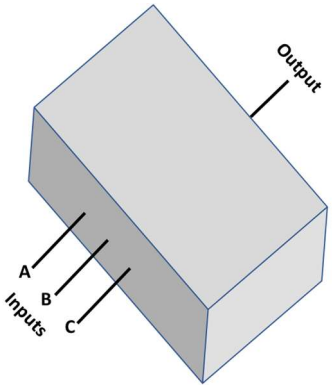


Figure 1. Black Box with Known Relationship between Inputs and Output

The following is based on an algorithm from the book *Logic and Boolean Algebra* [6]:

- For each row in a truth table that has a T in the right-hand column (i.e., a True output), do the following:
  - For each variable in the row that has the value F, take the negation of that variable
  - For each variable in the row that has the value T, take the variable as-is
  - Form the conjunction of the above terms.
- For each row in the truth table that has an F in the right-hand column (i.e., a False output), do nothing.
- Form the disjunction of the resulting conjunction for each row with T as an output.

To help clarify the above algorithm, we applied it to the truth table shown in Table 10.

Table 10. Truth Table with Unknown Associated Expression

A	B	C	Output
T	T	T	T
T	F	T	T
F	T	T	T
F	F	T	T
T	T	F	T
T	F	F	F
F	T	F	F
F	F	F	F

The entries in the extreme right-hand column of Table 11 are determined using the algorithm stated above.

Table 11 Determination of an Expression that Matches a Truth Table

A	B	C	Output	Expression Matching Input and Output for each Row whose Output is True
T	T	T	T	$A \wedge B \wedge C$
T	F	T	T	$A \wedge \neg B \wedge C$
F	T	T	T	$\neg A \wedge B \wedge C$
F	F	T	T	$\neg A \wedge \neg B \wedge C$
T	T	F	T	$A \wedge B \wedge \neg C$
T	F	F	F	-
F	T	F	F	-
F	F	F	F	-

The expressions in the right-hand column match input and output for a given row. For example, in the third row, the statement $\neg A \wedge B \wedge C$ is True if and only if A =False, B =True, and C =True.

When the disjunction of the expressions in the right-hand column is formed, the resulting expression (as shown immediately below) is True if and only if one of the outputs in the first five rows is True, and False otherwise:

$$(A \wedge B \wedge C) \vee (A \wedge \neg B \wedge C) \vee (\neg A \wedge B \wedge C) \vee (\neg A \wedge \neg B \wedge C) \vee (A \wedge B \wedge \neg C)$$

This gives us the desired result, i.e., an expression that matches the given truth table. However, this is not the simplest expression that generates the desired output. A simpler (in fact, the simplest) expression that generates the desire output is $(A \wedge B) \vee C$ (see Table 4).

4.9 Arguments

4.9.1 Definitions

An **argument** is a set of statements where one statement (the conclusion: Q) is affirmed on the basis of the others (the premises: $P_1, P_2, \dots, P_n$). The previous statement is represented in notation as $P_1, P_2, \dots, P_n \vdash Q$. The comma in the notation for an argument represents conjunction.

The argument $P_1, P_2, \dots, P_n \vdash Q$ is said to be **valid** if Q is true whenever all the premises are true; otherwise, the argument is said to be a **fallacy**.

For example, the **law of detachment** (defined as $A, A \Rightarrow B \vdash B$) is a valid argument. This law is also referred to as "**modus ponens**." The validity of this argument can be seen by consideration of the following truth table. Whenever the two premises are true (A and $A \Rightarrow B$ are true), then B is true, which is the case for the first row in the table.

A	B	$A \Rightarrow B$
T	**T**	**T**
T	F	F
F	T	T
F	F	T

Several other basic rules are as follows:

- **modus tollens**: $\neg B, A \Rightarrow B \vdash \neg A$
- **disjunctive syllogism**: $\neg A, A \vee B \vdash B$
- **hypothetical syllogism**: $A \Rightarrow B, B \Rightarrow C \vdash A \Rightarrow C$
- **constructive dilemma:** $A \vee B, A \Rightarrow C, B \Rightarrow D \vdash C \vee D$
- **simplification:** if $A \wedge B$ is true, one can infer A and B are true
- **conjunction:** if A and B are known to be true, one can infer $A \wedge B$ is true
- **addition:** If A is true, one can infer that $A \vee B$ is true

All of the above can be proven to be valid arguments. With use of the following theorem, we prove the hypothetical syllogism rule.

Theorem 4-2 The argument $P_1, P_2, \dots, P_n \vdash Q$ is valid if and only if the proposition $(P_1 \wedge P_2 \wedge \dots \wedge P_n) \Rightarrow Q$ is always true $(i.e., is\ a\ tautology)$.

Proof: The propositions $P_1, P_2, \dots, P_n$ are all true if and only if the proposition $P_1 \wedge P_2 \wedge \dots \wedge P_n$ is true. So, the argument $P_1, P_2, \dots, P_n \vdash Q$ is valid if and only if Q is true whenever $P_1 \wedge P_2 \wedge \dots \wedge P_n$ is true, i.e., $(P_1 \wedge P_2 \wedge \dots \wedge P_n) \Rightarrow Q$ is a tautology.

From Table 12, we see that $\big((A \Rightarrow B) \wedge (B \Rightarrow C)\big) \Rightarrow (A \Rightarrow C)$ is always true, and thus by Theorem 4-2 the hypothetical syllogism rule is valid.

Table 12. Truth Table for Law of Syllogism

A	B	C	$A \Rightarrow B$	$B \Rightarrow C$	$(A \Rightarrow B) \wedge (B \Rightarrow C)$	$A \Rightarrow C$	$\big((A \Rightarrow B) \wedge (B \Rightarrow C)\big) \Rightarrow (A \Rightarrow C)$
T	T	T	T	T	T	T	T
T	T	F	T	F	F	F	T
T	F	T	F	T	F	T	T
T	F	F	F	T	F	F	T
F	T	T	T	T	T	T	T
F	T	F	T	F	F	T	T
F	F	T	T	T	T	T	T
F	F	F	T	T	T	T	T

4.9.2 Conversion from Prose to Logical Arguments

Using the logic developed thus far, we can take a paragraph of prose, convert the prose to a logical argument and then determine the validity of the argument.

Various keywords are indicative of a premise, e.g., because, for, as, since, after all. These words are known as **premise indicators**. Other keywords indicate conclusions, e.g., therefore, so, thus, hence, accordingly. These words are known as **conclusion indicators**.

4.9.2.1 Modus Ponens Example

Professor Lopez's grade assignments are aligned with the guidelines from the university. So, the complaints about Professor Lopez's grade distributions from the students are not correct.

In this example, the word "so" indicates that the second sentence is the conclusion. One of the premises is implied, i.e., "If a professor's grade assignments are aligned with the university's guidelines, then students have no grounds for complaint." If we make the following assignments

A: a professor's grade assignments (distribution) are aligned with university guidelines

B: students have no grounds to complain about grade distributions

then the example can be written as $A, A \Rightarrow B \vdash B$, noting that Professor Lopez meets the criterion in premise A. This is exactly the modus ponens rule, which we already know to be a valid argument.

4.9.2.2 Conjunction and Modus Ponens Example

In order to be good at basketball, one needs to practice shooting, do resistance and aerobic training, and study the various defensive and offensive strategies of the sport. Sally practices shooting, and does engage in resistance and aerobic training but does not know well the defensive and offensive strategies of basketball. Thus, Sally is not a good basketball player at this point.

The "in order to" phrase indicates the three premises that follow in the first sentence. The second sentence of the example is another premise. The conclusion is indicated by the word "thus." More formally, we have

A: Practice shooting

B: Do resistance and aerobic training

C: Study the various defensive and offensive strategies of basketball

D: Be good at basketball

The proposition $A \wedge B \wedge C \Rightarrow D$ is true.

The wording in the informal argument statement implies that if A, B or C is missing, then not being good at basketball is the outcome, i.e., $\neg D$. In the case of Sally, we have the proposition $A \wedge B \wedge \neg C \Rightarrow \neg D$.

So, the prose description can be written more formally as the argument

$$A, B, \neg C, (A \wedge B \wedge \neg C) \Rightarrow \neg D \vdash \neg D$$

The above argument will be proven to be valid in the next section.

4.9.2.3 College Entrance Example

If a high school student studies hard, he or she will get good grades. If a high school student gets good grades and does well on the college entrance exams, he or she will get into a good college.

Albert (a high school student) has studied hard and has done well in college entrance exams. Therefore, Albert will get into a good college.

We make the following assignments:

A: a high school student who studies hard

B: a high school student who gets good grades

C: a high school student who does well on college entrance exams

D: a high school student who gets into a good college.

Noting that Albert meets the criteria in premises A and C, we can recast the prose statements concerning Albert in the following argument:

$$A, C, A \Rightarrow B, B \wedge C \Rightarrow D \vdash D$$

The above will be proven as valid in the next section.

4.9.3 Proofs

Logic arguments require proof of validity. In Section 4.9.1, a truth table was used to prove a variation of the hypothetical syllogism rule. Another approach is to systematically use basic rules (some of which were stated in the Section 4.9.1) to prove more complex arguments as being valid.

The general approach of a proof is to list the premises and then derive intermediate results until the conclusion is reached.

The proof of validity for the argument concerning the basketball example in Section 4.9.2.2 goes as follows:

1. A (premise)
2. B (premise)
3. $\neg C$ (premise)
4. $(A \wedge B \wedge \neg C) \Rightarrow \neg D$ (premise)
5. $A \wedge B \wedge \neg C$ (conjunction applied to lines 1, 2 and 3)

6. $\neg D$ (modus ponens applied to lines 4 and 5). Thus, the argument is valid.

The proof of validity for the argument concerning the college entrance example in Section 4.9.2.3 goes as follows:

1. A (premise)
2. C (premise)
3. $A \Rightarrow B$ (premise)
4. $B \wedge C \Rightarrow D$ (premise)
5. B (modus ponens applied to lines 1 and 3)
6. $B \wedge C$ (conjunction applied to lines 2 and 5)
7. D (modus ponens applied to lines 4 and 6). Thus, the conclusion is valid.

4.9.4 Equivalence Rules

The equivalence rules, in the list below, allow for one expression to be replaced by another and vice versa. Such rules are needed for the validity proof of more complex arguments.

Recall that two expressions are equivalent if they have identical truth tables and thus, the following rules can be proven by comparing the truth tables of each pair of expressions:

- **double negation:** $\neg(\neg A)$ is equivalent to A
- **commutation:** (see the commutativity laws in Theorem 4-1)
- **redundancy:** $A \vee A, A \wedge A$ and A are equivalent
- **contraposition:** $A \Rightarrow B$ is equivalent to $\neg B \Rightarrow \neg A$
- **association:** (see the associativity laws in Theorem 4-1)
- **exportation**: $(A \wedge B) \Rightarrow C$ is equivalent to $A \Rightarrow (B \Rightarrow C)$
- **material implication**: $A \Rightarrow B$ is equivalent to $\neg A \vee B$
- **distributive laws:** (see the distributive laws in Theorem 4-1)
- **De Morgan's laws:** (see Theorem 4-1)
- **material equivalence:** $A \Leftrightarrow B$ is equivalent to $(A \Rightarrow B) \wedge (B \Rightarrow A)$.

As an example of using one of the rules from the above list, we prove the validity of the argument $\neg A \vee B, B \Rightarrow A \vdash A \Leftrightarrow B$ as follows:

1. $\neg A \vee B$ (premise)
2. $B \Rightarrow A$ (premise)
3. $A \Rightarrow B$ (material implication applied to line 1)
4. $A \Leftrightarrow B$ (material equivalence applied to lines 2 and 3). Thus, the conclusion is valid.

As a second example, consider the argument $(A \wedge B) \vee (A \wedge B), A \Rightarrow (B \Rightarrow C) \vdash C$. The validity proof is as follows:

1. $(A \wedge B) \vee (A \wedge B)$ (premise)
2. $A \Rightarrow (B \Rightarrow C)$ (premise)
3. $(A \wedge B) \Rightarrow C$ (exportation applied to line 2)
4. $A \wedge B$ (redundancy applied to line 1)
5. C (modus ponens applied to lines 3 and 4). Thus, the conclusion is valid.

4.10 Exercises

1. Write an equivalent statement for A NOR B using only NAND operations? ***Hint****: See the Wikipedia article "NAND Logic" [7].*
2. What is the truth table for the statement $(A \Rightarrow B) \Rightarrow C$?
3. Show that $A \Rightarrow B$ and $\neg B \Rightarrow \neg A$ have the same truth tables and are thus equivalent. [**Remark**: This fact is commonly used in many mathematical proofs since in some cases, it is easier to prove that the negation of one statement implies the negation of another.]
4. Show that $(A \Rightarrow B) = (\neg A) \vee B$.
5. Prove the first distributive law in Theorem 4-1.
6. Prove that $(A \Rightarrow B) \vee (\neg B) = \neg A$. ***Hint****: Write down the truth table for the expression, showing the outcome is always true. Another approach is to use the result of Exercise #4 above along with the distributive and commutative laws. To prove simple expressions, it is typically easier to write down the truth table. For more complex expressions, it is typically easier to reduce the expression using simpler (already known) expressions rather than writing down a large truth table.*
7. Prove that $(A \Leftrightarrow B) = \big((A \Rightarrow B) \wedge (B \Rightarrow A)\big)$. ***Hint****: write down the truth table for each side of the equation.*
8. Write a logical expression for the following statement about baseball: "A batter is out if he or she has three strikes, hits the ball in the air and the ball is caught before it hits the ground, or hits the ball on the ground but is thrown out at first base." ***Hint****: divide the statement into smaller statements, e.g., A: "the batter is out" and B: "he has three strikes", and then connect the smaller statements using logical operators.*
9. If $A \wedge B = A \wedge C$ is it true, is the proposition $B = C$ also true? ***Hint****: Use a truth table to show that the two expressions are not the same for all input combinations of A, B and C.*

10. Find an expression that has the same output as the following truth table for the given value combinations of A, B and C:

A	B	C	$Output$
T	T	T	T
T	F	T	F
F	T	T	T
F	F	T	F
T	T	F	T
T	F	F	F
F	T	F	T
F	F	F	F

11. Prove modus tollens is valid by consideration of the associated truth table.
12. Prove the exportation rule by comparing the truth table of both expressions.
13. Prove the validity of the argument $\neg(A \vee B), C \vdash (C \wedge \neg A) \wedge \neg B$.

5 First-Order (or Predicate) Logic

5.1 Overview

First-order logic is an extension of propositional logic that allows for propositions with variables. For example, the proposition "x is an even number" (where x is a variable) is allowed in first-order logic but not supported in propositional logic. Propositions with a variable (or variables) are represented with a letter and the associated variable (or variables). Some examples:

- The proposition "x is an even number" is represented as $A(x)$ or alternately Ax. For each value of x, $A(x)$ is a proposition that is either true or false, e.g., $A(3)$ is false since 3 is an odd number. To be clear, $A(x)$ is an example of a first-order proposition.
- The statement "the integer x divides the integer y exactly, i.e., with zero remainder" is a first-order proposition. It can be represented as $D(x, y)$. $D(2,8)$ is true since 2 divides 8 exactly 4 times.

Further, there are two types of quantifiers that are used in conjunction with first-order logic:

- Let $A(x)$ be a proposition with variable x where x is a member of a given set S (e.g., the set of integers). One can write the expression $(\forall x \in S)A(x)$ which reads as "for every x in set S, the proposition $A(x)$ is true." The symbol ∀ means "for every" and it is known as the **universal quantifier**.
    - For example, "for every element x in the set of even numbers E, the number x is divisible by 2" can be written as $(\forall x \in E)A(x)$ where $A(x)$ is the proposition "x is divisible by 2."
    - The negation of the universal quantifier means "not for every" and can be represented as (¬∀) or as an upside-down A with a slash through it (but this is not common in most fonts and will not be used in this book).
- Defining $A(x)$ and S as above, one can also make statements such as $(\exists x \in S)A(x)$ which reads as "there exists an element of S such that the proposition $A(x)$ is true." The symbol ∃ means "there exists" and is known as the **existential quantifier**.
    - For example, "there exists an integer value of x such that $x - 7 = 10$" can be written as $(\exists x \in \mathbb{Z})F(x)$ where $\mathbb{Z}$ represents the set of integers and $F(x)$ is the proposition "$x - 7 = 10$."
    - The symbol for the negation of the existential quantifier is ∄ which means "there does not exist." Alternately, one can write (¬∃).

Again, it should be emphasized that first-order logic is an extension of propositional logic. For a given value of x, the proposition $A(x)$ is just like any of the other propositions that were covered in the previous section on propositional logic (all the various relationships and laws hold true).

5.2 Provable Identities

First-order logic (with the noted additions of variables in propositions, the universal quantifier and the existential quantifier) has additional properties beyond those in propositional logic. A small subset of the additional properties is listed in the following theorem.

Theorem 5-1 The following properties hold true for first-order logic:

- $\neg\forall x P(x) \Leftrightarrow \exists x \neg P(x)$ (a version of De Morgan's Law)
    - It is helpful to read the above statement out loud. The left-hand side says "it is not the case that for every $x, P(x)$ holds true." The right-hand side says "there exists at least one x such that $P(x)$ is false."
    - This can also be written as $\forall x P(x) \Leftrightarrow \neg[\exists x \neg P(x)]$.
- $\nexists x P(x) \Leftrightarrow \forall x \neg P(x)$ (yet another variant of De Morgan's Law)
    - This can also be written as $\exists x P(x) \Leftrightarrow \neg[\forall x \neg P(x)]$.
- $\forall x P(x) \wedge \forall x Q(x) \Leftrightarrow \forall x Q(x) \wedge \forall P(x)$ (Commutative property for conjunction relative to the universal quantifier)
- $\exists x P(x) \vee \exists x Q(x) \Leftrightarrow \exists x Q(x) \vee \exists P(x)$ (Commutative property for disjunction relative to the existential quantifier)

Proof: Only proofs of De Morgan Laws are provided.

For the first De Morgan Law, the proof comes in two parts.

First, show that $\neg\forall x P(x) \Rightarrow \exists x \neg P(x)$.

> $\neg\forall x P(x)$ means that $P(x)$ is not true for every value of x. So, there is some value x_0 such that $P(x_0)$ is false, i.e., $\neg P(x_0)$ is true and thus $\exists x \neg P(x)$.

Going in the other direction, show that $\exists x \neg P(x) \Rightarrow \neg\forall x P(x)$.

> $\exists x \neg P(x)$ means there is some value of x (say x_0) such that $\neg P(x_0)$ is true and thus $P(x_0)$ is false. But $P(x_0)$ being false implies $P(x)$ is not true for every value of x, i.e., $\neg\forall x P(x)$.

The proof of the second De Morgan Law is also in two parts.

First, show that $\nexists x P(x) \Rightarrow \forall x \neg P(x)$.

> $\nexists x P(x)$ means there is no x such that $P(x)$ is true, i.e., $P(x)$ is false for every value of x or in other words $\neg P(x)$ is true for every value of x. Thus, $\forall x \neg P(x)$.

Going the other direction, show that $\forall x \neg P(x) \Rightarrow \nexists x P(x)$.

> $\forall x \neg P(x)$ means that $\neg P(x)$ is true for every value of x which implies that $P(x)$ is false for every value of x, i.e., there is no x such that $P(x)$ is true. Thus, $\nexists x P(x)$ ■

5.3 Scope of a Quantifier

The **scope of a quantifier** (either universal or existential) is the first complete expression that follows the quantifier. Propositions, such as $P(x)$ or $Q(x)$, are the smallest complete expressions and serve as the building blocks for more complex expressions. In the proposition $\forall x P(x)$, $P(x)$ is the first complete expression after the quantifier and thus, falls within the scope of the quantifier. In the proposition $\forall x P(x) \Rightarrow Q(y)$, $P(x)$ is within the scope of the quantifier $\forall x$, but $Q(y)$ is not since $P(x)$ is the first complete expression after the quantifier $\forall x$. However, in the expression $\forall x(P(x) \Rightarrow Q(y))$, $Q(y)$ is within the scope of the quantifier $\forall x$.

A **bound variable** is one that falls within the scope of its "own" quantifier, i.e., a quantifier using the same variable. A **free variable** is one that does not fall within the scope of its own quantifier. A variable is free if and only if it is not bound.

For example, in the proposition $\forall x(P(x) \Rightarrow Q(y))$, the x in $P(x)$ is bound, but the y in $G(y)$ is free since, although y does fall within the scope of the quantifier, it is not a matching quantifier (i.e., x is being quantified not y). In the proposition $\forall x P(x) \Rightarrow Q(x)$, the x in $P(x)$ is bound, but the x in $Q(x)$ is free, since $Q(x)$ falls outside the scope of the quantifier. However, in the proposition $\forall x(P(x) \Rightarrow Q(x))$, all occurrences of x are bound.

5.4 Arguments

Section 0 covered arguments for propositional logic. Arguments can be extended to first order logic if several rules are added (as listed below). In what follows, the shortened notation for $P(x)$ is used, i.e., Px.

- ∀-elimination (or universal instantiation) – this allows for the elimination of a universal quantifier and replacement of the variable with a constant. For example, consider the expression $\forall x((Px \wedge Qx) \Rightarrow Rx)$. The ∀-elimination rule applies to the entire statement and allows one to replace a quantified variable with either a free variable or by any specific constant. For example, in the statement above, we can replace x by the constant a to get $Pa \wedge Qa \Rightarrow Ra$.
- ∃-introduction (or existential generalization) – this allows for a statement that holds true for a given constant, to be replaced by a statement with an existential quantifier. For example, $Pa \wedge Qa \Rightarrow Ra$ can be replaced by $\exists x((Px \wedge Qx) \Rightarrow R(x))$.
- Existential introduction (or existential instantiation) – allows one to remove an existential quantifier and replace it with a free variable, but only with a free variable that appears free nowhere earlier in a given validity proof. For example, the statement $\exists x Ax$ can be replaced by Ay as long as the free variable y does not appear earlier in a validity proof.
- Universal introduction (or universal generalization) – if a statement can be proven true for any arbitrary constant, then it must be true for all things. This rule allows one to move from a particular statement about an arbitrary object to a general statement using a universal quantifier. More formally:
    - If one can prove $S \vdash P(c)$, then $S \vdash \forall x Px$ is true, where S is a set of propositions and c is an arbitrary constant that is not present in any of the propositions of S.

There are additional (and more complex) rules for the reduction of first order logic arguments, see the section entitled "Expanded Proof Method with Predicates and Quantifiers" in the book by Byerly [8].

5.4.1 Single Variable Example

As an example of applying the above rules, consider the following argument:

> Ted is a ship pilot, but Ted does not like Donna the mountain climber. If anyone gets motion sickness, he or she is not a pilot. So, there is someone who does not get motion sickness and who does not like Donna.

Make the following assignments:

- Let P stand for the property of being a ship pilot.
- Let Q stand for the property of liking Donna the mountain climber.
- Let R stand for the property of getting motion sickness
- Let α stand for Ted.

The above argument can be written as

$$\forall x(Rx \Rightarrow \neg Px), P\alpha \wedge \neg Q\alpha \vdash \exists x(\neg Rx \wedge \neg Qx)$$

The validity proof goes as follows:

1. $\forall x(Rx \Rightarrow \neg Px)$ (premise)
2. $P\alpha \wedge \neg Q\alpha$ (premise)
3. $R\alpha \Rightarrow \neg P\alpha$ (∀-elimination applied to line 1)
4. $P\alpha$ (simplification rule applied to line 2)
5. $\neg(\neg P\alpha)$ (double negation applied to line 4)
6. $\neg R\alpha$ (modus tollens applied to lines 3 and 5)
7. $\neg Q\alpha$ (simplification rule applied to line 2)
8. $\neg R\alpha \wedge \neg Q\alpha$ (conjunction applied to lines 6 and 7)
9. $\exists x(\neg Rx \wedge \neg Qx)$ (∃-introduction applied to line 8). Thus, the argument is valid.

5.4.2 Multi-variable Example

For a multi-variable example, consider the following variation of the example from Section 4.9.2.1:

> If a professor's grade distribution for a class is aligned with the university's guidelines then a student in that class has no grounds to complain about grade distributions. For class α, Professor Lopez's grade distribution is aligned with the university's guidelines. Jane is a student in Professor Lopez's class α and thus, Jane has no ground to complain about Professor Lopez's grade distribution for class α.

Make the following assignments:

- Let Pxz be the property that the grade distributions of Professor x, for class z, is aligned with university guidelines.
- Let Qyz be the property that student y is in class z
- Let Ryz be the property that Student y has no grounds to complain about the grade distribution for class z
- Let constant l represent Professor Lopez and constant j represent Jane.

The above argument can be written as

$$\forall xyz((Pxz \wedge Qyz) \Rightarrow Ryz), Pl\alpha, Qj\alpha \vdash Rj\alpha$$

The validity proof goes as follows:

1. $\forall xyz((Pxz \wedge Qyz) \Rightarrow Ryz)$ (premise)
2. $Pl\alpha$ (premise)
3. $Qj\alpha$ (premise)
4. $(Pl\alpha \wedge Qj\alpha) \Rightarrow Rj\alpha$ (by applying ∀-elimination to variables x, y and z in line 1)
5. $Pl\alpha \wedge Qj\alpha$ (conjunction of lines 2 and 3)
6. $Rj\alpha$ (modus ponens applied to lines 4 and 5). Thus, the argument is valid.

5.5 Paradoxes

The Wikipedia article on paradoxes [11] provides the following definition:

> "A paradox is a logical statement that seems to contradict itself. It is a statement that, despite apparently valid reasoning from true premises, leads to an apparently-self-contradictory or logically unacceptable conclusion. A paradox involves contradictory-yet-interrelated elements that exist simultaneously and persist over time."

The website WikiDiff provides the following definition of "paradox":

> A self-contradictory statement, which can only be true if it is false, and vice versa.

However, W. V. Quine (1962) distinguished between three kinds of paradoxes [12]:

- A **veridical paradox** appears to be absurd but upon further study can be proven to be true, e.g., see the Drinking Paradox below.
- A **falsidical paradox** appears to be false and in fact can be proven to be false. An example of this is Zeno Paradox of Motion (which we cover in Section 6.6.6).
- A paradox that is in neither of the classes described above is called an **antinomy**, which reaches a self-contradictory result by properly applying accepted ways of reasoning, e.g., see Berry's paradox and the liar's paradox below.

There is a large collection of logical paradoxes that have been debated by professional logicians, in some cases for centuries. A few of these paradoxes are discussed in the subsections below.

5.5.1 Berry's Paradox

Berry's paradox has several versions. The following is one of the more basic versions:

> What is "the smallest integer that cannot be expressed in less than thirteen words." Since the quoted expression in the previous sentence has 12 words, to which set does the integer it describes (call it x) belong: the set of integers that can be expressed with less than 13 words (call it set A), or the set of integers that can only be expressed with 13 words or more (call it set B)?

If we say x is in A, then the sentence that describes x is false. If we say x is in B, then we have a contradiction since the sentence that describes x has only 12 words. Either answer leads to a contradiction. Thus, Berry's paradox is an antinomy.

The problem arises from the ambiguous use of the term "expressed." We can fix the statement (i.e., remove the paradox) if rather than "expressed" we say "represented as a binary number using

"one" for the binary digit 1 and "zero" for the binary digit 0. For the sake of brevity, while still making the point, let's just say "less than 4 words." From Table 13, we can see that 8 is the smallest integer that cannot be expressed in less than 4 words (using a combination of "one" and "zero" that match the binary expression for the integer).

Table 13. Fixing Berry's Paradox

Decimal Number	Binary representation	Expression in words
0	0	zero
1	1	one
2	10	one zero
3	11	one one
4	100	one zero zero
5	101	one zero one
6	110	one one zero
7	111	one one one
8	1000	one zero zero zero

5.5.2 The Liar's Paradox

In its simplest form, the liar's paradox goes as follows:

This sentence is false.

The question is "whether the above sentence is true or false?" If one claims the sentence is false, then what the sentence claims is true and thus we have a contradiction. If one claims the sentence is true, then the (assumed true statement) claims that it is false and again, we have a contradiction. So, the statement is neither true nor false, and we have an antinomy.

There are many versions of this paradox. Some versions involve several statements that collectively lead to a contradiction, e.g.,

A: The following statement is true.

B: The preceding statement is false.

If A is true, then statement B leads to a contraction. If A is false, then the implication is that B is false which, in turn, implies statement A is true, another contradiction.

5.5.3 Drinking Paradox

In his book "What is the Name of this Book" [13], logician Raymond Smullyan tells the story about a man who buys everyone a drink when he has a drink at a particular bar. This man makes the statement "when I drink, everybody drinks." Smullyan asks the reader "does there really exist someone such that if he drinks, everybody drinks?" As opposed to the bar story, usage of the term "drink" in the question posed by Smullyan means "drink alcohol in general and not at a specific time and place (as was implied in the bar story)."

Let $D(x)$ be the proposition "x drinks" and P be the set of all people. The statement to be proved is

$$(\exists x \in P)\big(D(x) \Rightarrow (\forall y \in P)D(y)\big)$$

[The above statement is provided just to give the reader some additional practice with formal notation. The explanation below does not make use of this formal statement.]

The solution goes back to a point made earlier in this document, i.e., anything can be implied by a false statement.

There are two cases, i.e., either everybody drinks or not.

- Case 1: Everybody drinks. Take any person (Fred). Since it was assumed that everybody drinks then clearly, Fred drinks. Thus, it is true that if Fred drinks, then everybody drinks. Thus, there is at least one person (Fred) such that if he drinks then everybody drinks.
    - The claim that the statement "if Fred drinks, then everybody drinks" may seem odd since that fact that Fred drinks is not the cause of everybody drinking. Another example may help. Consider the statement "if $2 + 2 = 4$, then $\frac{24}{6} = 4$". This statement is true. It may help to write the statement more formally. Let A be the statement $2 + 2 = 4$ and B be the statement $\frac{24}{6} = 4$. Then we can write the statement as $A \Rightarrow B$ which is true since both A and B are true (see Table 6).
- Case 2: Not everybody drinks and so there is at least one person who doesn't drink (Alice). Since it is false that Alice drinks, then it is true that **if** Alice drinks, everybody drinks (this statement is vacuously true). For this case, there exists a person (Alice) such that if she drinks, everybody drinks. In formal notation, let A be the statement "Alice drinks" and B be the statement "everybody drinks". Then the statement $A \Rightarrow B$ is true since A is false and B is true (see Table 6).

So, in either case, it follows that there does exist someone such that if he or she drinks, then everybody drinks.

5.5.4 The Unexpected Hanging

The paradox of the unexpected hanging goes as follows:

> A judge tells a convicted criminal: "You are sentenced to hang at noon on a day in the following week but that the execution will be a surprise to you, i.e., you will not know the day of the hanging until the executioner comes to retrieve you shortly before noon on the day of the execution."

[For this discussion, assume the week starts on Sunday and ends on the following Saturday.]

The criminal reasons that the execution cannot happen on Saturday since after 12 noon has passed on Friday, he will know that he is to be hanged on Saturday and thus no surprise. Next, he reasons that the execution cannot happen on Friday after 12 noon has passed on Thursday, since he will know that he is to be hanged on Friday (having already reasoned that Saturday is not possible) and thus no surprise. Similarly, the criminal reasons that he cannot be executed on any of the days specified by the judge and thus thinks he will not be executed. However, Wednesday comes and the criminal is hanged, and it is unexpected to the criminal based on his reasoning.

Is the statement from the judge true?

This paradox has generated hundreds of papers in mathematics and philosophy journals. In his book "The Unexpected Hanging and Other Mathematical Diversions" [9], Martin Gardner explains the paradox from various perspectives and describes several equivalent paradoxes. In his technical

paper "The Surprise Examination or Unexpected Hanging Paradox" [10], Timothy Chow clarifies and reformulates the problem as a self-referential statement, e.g., "You will be executed sometime during the next week and the date will not be deducible in advance using this statement as an axiom." The term "axiom" is meant to be the assumption under which the criminal is to reason concerning which day he will be executed. In his paper, Chow goes on to prove that the self-referential statement is false. So, as reformulated, the judge's statement to the criminal is false. Not all agree to this solution and the debate continues.

[Author's Remark: Analysis of this problem provides a good thinking exercise. However, from a practical point of view, the statement is fundamentally unclear. If I were the judge and legally bound to my statement, I would have said something like "I will make arrangements with the prison authorities and request that, within the next week, you be hung by the neck until dead."**]**

5.6 Exercises

1. If $Walk(x)$ is the proposition "x walks", $Talk(x)$ is the proposition "x talks" and H is the set of humans, what do the following mean:
   - $\forall x \in P(Walk(x) \vee Talk\ (x))$
   - $\exists x \in P(Walk(x) \vee Talk\ (x))$
2. Let $P(x)$ be the proposition "Steve likes to eat x." Write the following phases in formal notation:
   - Steve likes to eat ice cream.
   - Steve likes to eat at least one thing. ***Hint****: use the universal quantifier. The relevant set is the set of all edible things (call it E).*
   - Steve likes to eat everything. ***Hint****: use the existential quantifier.*
3. As illustrated in the example of Section 5.4.2, it is possible to have propositions with more than one variable. For example, in the previous exercise, we could substitute a variable for Steve and get a more general proposition, i.e., $P(y, x) =$ "y likes to eat x" where y is a member of the set of humans and x is a member of the set of edible things. For example, $P(Bob, pickles)$ means that "Bob likes to eat pickles." Write the following phrases in formal notation:
   - Alice likes to eat popcorn.
   - At least one person likes to eat everything. ***Hint****: apply the existential quantifier to y and the universal quantifier to x.*
   - Everybody likes to eat everything. ***Hint****: apply the universal quantifier to both x and y.*
4. What is the negation of $\exists x \forall y P(x, y)$? ***Answer****: Use De Morgan's laws several times in the following sequence of equivalences:*

$$\neg\big(\exists x \forall y P(x, y)\big) = \neg\exists x\ \big(\forall y P(x, y)\big) = \forall x \neg(\forall y\big(P(x, y)\big) = \forall x \exists y \neg P(x, y)$$

5. Show that the negation of $\forall x \forall y P(x, y)$ is $\exists x \exists y \neg P(x, y)$? ***Hint****: Similar to the solution of Exercise #4*.
6. Prove that the following argument is valid:

$\forall x(Ax \Longrightarrow (Bx \vee Cx)), \exists x(Ax \wedge \neg Cx) \vdash \exists x(Ax \wedge Bx)$ ***Hint**: Use existential instantiation, universal instantiation, several other rules and then existential generalization in the last step.*

7. Is the following list of statements a paradox?
    - One plus one is two.
    - One plus one is three.
    - Exactly one sentence in this list is true.

6 Sets

6.1 Overview

Set theory is a branch of mathematics that entails the study of sets, i.e., collections of objects. The objects in a set can be anything, e.g., books, cars, numbers. A set can be composed of objects of different types or objects of the same type. Consider the following examples:

- A = {1, 2, 3, 4}
    - Set A includes a limited set of objects of the same type (i.e., the set of the first four positive integers).
- B = {green, 12, “US Air Force 1”, ¼, “Stephen Fratini”}
    - Set B has a mixture of different types. Further, some of the types are abstract (such as “green”) and some are specific concrete things (such as “US Air Force 1” and “Stephen Fratini”).
- C = {1, 2, 3, 4, …}
    - Set C is the infinite set of positive integers. The “…” indicates that the pattern continues indefinitely.
- D = {A, B, C} where A, B and C are as defined above.
    - A set can be composed of other sets.
- Some sets are described in words but not explicitly enumerated, e.g., the set of grains of sand on the planet Earth.

6.2 Terminology

Two sets are **equal** if and only if they have the same elements (the order of the elements does not matter). Equivalently, sets A and B are equal if every element of A is in B, and every element of B is in A. The latter statement is typically used to prove two sets are equal.

- For example, if $A = \{x, 1, y, 2, z, 3\}$ and $B = \{1, 2, 3, x, y, z\}$ then A and B are equal (written as $A = B$).
- Given the definition of equal sets, it follows that repeated elements are irrelevant and are typically not even noted. For example, $A = \{1, 2, 2, 3, 4\}$ is equal to $B = \{1,2,3, 4\}$ since every element in A is in B and every element in B is in A.

The contents of a set are called **elements**. In terms of notation, the expression “x is an element of set A” is written as $x \in A$. There is a special set with no elements known as the **empty set**, which is written as either { } or ϕ. A **universal set** (Ω) is the set of all objects within a domain of interest. For example, if the domain of interest is the integers (all positive and negative whole numbers, and zero) then the universal set would be the integers. A universal set which contains all objects (including itself) leads to what is called Russell's paradox and is consequently not allowed in most formulations of set theory [14]. So, in this book, the focus is only on universal sets for a specific domain of interest.

Sets can be combined to form another set (this is commonly referred to as the **union** of sets). For example, if $A = \{ 1, 2, 3 \}$ and $B = \{ a, b, c \}$ then the combination (union) of the two is

$\{ 1, 2, 3, a, b, c \}$. In terms of notation, the expression "the union of sets A and B" is written as $A \cup B$.

The **intersection** of two sets A and B, denoted by $A \cap B$, is the set containing all elements of A that also belong to B (or equivalently, all elements of B that also belong to A). For example, if $A = \{ green, blue, red, orange \}$ and $B = \{ purple, orange, blue \}$ then the intersection of A and B is $\{ orange, blue \}$.

Set union and set intersection are similar in concept to disjunction and conjunction, respectively. These similarities will be discussed further in Section 7.3 regarding Boolean algebra.

Set B is said to be a **subset** of A if every element of B is also an element of A. The shorthand notation for this is $B \subset A$ (this typically implies that B cannot be equal to A, sometimes stated as "B is a strict subset of A"). If B is a subset of A which might also be A itself, one writes $B \subseteq A$. If B is not a subset of A, this is written as $B \not\subset A$.

Set difference is defined as the elements in a given set but not in another set. This can be written as $A - B = \{ x \in A \mid x \notin B \}$ which is read as A – B is the set of elements x in A such that x is not an element of B. The vertical bar in the notation means "such that" (sometimes a colon is used instead of a vertical bar). There are several cases to consider regarding the relationship of A and B when computing a set difference:

- $B \subset A$: For example, $A = \{ a, b, c, \ldots, x, y, z \}$ (i.e., A is the entire alphabet) and $B = \{ x, y, z \}$ which implies $A - B = \{ a, b, c, \ldots, u, v, w \}$.
- $B \not\subset A \; but \; A \cap B \neq \phi$ (i.e., B is not a subset of A, but the intersection of A and B is not empty): For example, $A = \{ a, b, c, d, e, f \}$ and $B = \{ d, e, f, g, h, i \}$ which implies $A - B = \{ a, b, c \}$.
- $A \cap B = \phi$: For example, $A = \{ a, b, c \}$ and $B = \{ d, e, f, g, h, i \}$ which implies $A - B = A$.

When the domain of interest is known (e.g., the positive integers or the alphabet), one can talk about the **complement** of a given set with respect to the domain of interest. For example, if the set of positive integers is the domain of interest and A is the set of odd numbers (i.e., $A = \{1,3,5, \ldots\}$), the complement of A is the set of even numbers which is written as $\neg A = \{ 2, 4, 6, \ldots \}$. This notation is essentially a shorthand for the equivalent statement $\Omega - A$, i.e., $\neg A = \Omega - A$ where Ω is the entire domain of interest. Further, $A - B$ is equivalent to $A \cap \neg B$.

For finite sets, the **cardinality** is the number of elements in a set. For example, the cardinality of the set $A = \{ a, b, c, \ldots, z \}$ is 26 (in reference to the alphabet used by English speakers). For infinite sets, it turns out that there are several (actually, an infinite number) of cardinalities. The smallest infinite cardinality is call **countable**, and defined as any set that can be mapped in a one-to-one manner with the set of positive integers (denoted as $\mathbb{N}$). For example, the set of all integers $\mathbb{Z} = \{ \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots \}$ is countable. To see this, consider the mapping of $\mathbb{Z}$ *to* $\mathbb{N}$:

Table 14. Mapping of $\mathbb{Z}$ to $\mathbb{N}$

…	-4	-3	-2	-1	0	1	2	3	…
…	7	5	3	1	2	4	6	8	…

In the above mapping, every element of $\mathbb{Z}$ is mapped to one and only one element of $\mathbb{N}$, and vice versa. For example, $-42 \in \mathbb{Z}$ is mapped to $(-2 \times -42) - 1 = 83 \in \mathbb{N}$ and $97 \in \mathbb{Z}$ is mapped to

$(97 + 1) \times 2 = 49 \in \mathbb{N}$. In general, a negative integer n is mapped $-2n - 1$, and a non-negative integer n is mapped to $2n + 2$.

For a thorough discussion of infinite set cardinalities, see the Wikipedia article on Cardinality [16].

6.3 Venn Diagrams

A Venn diagram is a 2-dimensional depiction of the containment relationships among 2 or more sets. In 1880, John Venn proposed what he called "Eulerian Circles" as a way to represent logical propositions [15]. Eulerian circles were later renamed as "Venn diagrams."

For example, Figure 2 depicts the intersection of sets A, B and C. Each of the three sets (A, B and C) are represented as circles and the common intersection is the darker-shaded, triangular-shaped region in the center. The various elements are shown within their containing sets. By examining the figure, one can determine that

- A = {1, 2, 3, 11, 12, 13, 14, 17}
- B = {4, 5, 6, 11, 12, 15, 16, 17}
- C = {7, 8, 9, 13, 14, 15, 16, 17}.

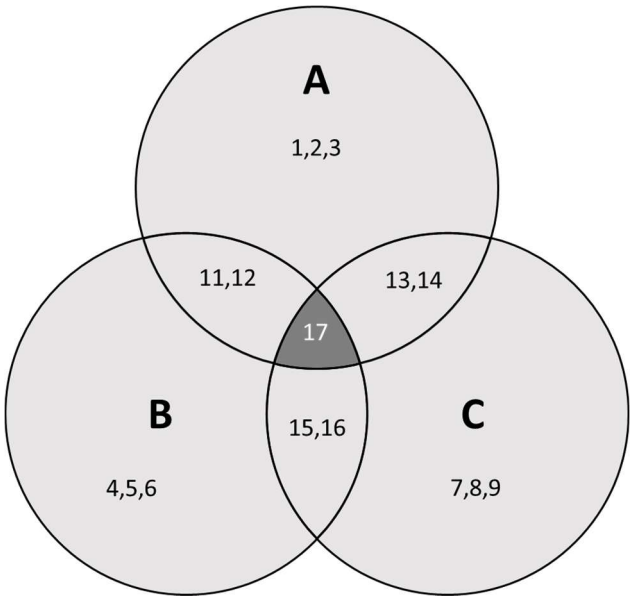


Figure 2. Venn diagram showing the intersection of three sets

6.4 Theorems

Several theorems are listed here for usage in other parts of this book. For instructive purposes, proofs are provided for some of the theorems.

Theorem 6-1 For every set $A, \emptyset \subset A$.

Proof: By definition, a set is a subset of another if every element of that set is an element of the containing set. However, ϕ has no elements and so the statement to be proved is vacuously true.

The above is a direct proof. An indirect proof (by way of contradiction) is also possible. Assume the opposite of the statement to be proved is false, i.e., ϕ is not a subset of A (written as $\phi \not\subset A$). If ϕ is not a subset of A, then ϕ must have an element that is not in A. However, ϕ has no elements and so there is a contradiction. Thus, the assumption must be false and in fact, ϕ is a subset of A■

In addition to being instructive from a thought process point of view, the above proof also forces one to clearly understand the definition of set equality. If, for example, the definition of set equality was modified to read "A is a subset of B, if every element of A is an element of B and A has a least one element" then the above theorem would not be true, even though, on the surface, the modified definition looks reasonable.

Theorem 6-2 If $A \subset B$ and $B \subset A$, then $A = B$.

Proof: This is almost a restatement of the definition of set equality. If $x \in A$ then $A \subset B$ implies $x \in B$. Similarly, if $x \in B$ then $B \subset A$ implies $x \in A$■

Theorem 6-3 The following properties hold true concerning set union:

a. $A \cup \phi = A\ (identity)$
b. $A \cup B = B \cup A\ (commutative)$
c. $A \cup (B \cup C) = (A \cup B) \cup C\ (associative)$
d. $A \cup A = A\ (idempotent)$
e. $A \subset B\ if\ and\ only\ if\ A \cup B = B.$

Proof: The proofs of the identity, commutative and idempotent properties are left to the reader. All three of these proofs follow the pattern used to prove the associative property which follows below.

To prove the associative property, we take an element of the left-side of the equation and show it is also an element of the right-side, and vice versa. Take any $x \in A \cup (B \cup C)$, then $x \in A\ or\ x \in B \cup C$.

- If $x \in A$ then $x \in A \cup B$ which implies that $x \in (A \cup B) \cup C$.
- If, on the other hand, $x \in B \cup C$ then $x \in B\ or\ x \in C$. This leads to the following two sub-cases:
    - If $x \in B$, then $x \in A \cup B$ which implies $x \in (A \cup B) \cup C$.
    - If $x \in C$, then $x \in (A \cup B) \cup C$.

So, in all cases, if $x \in A \cup (B \cup C)$ then $x \in (A \cup B) \cup C$. This just shows that $A \cup (B \cup C) \subset (A \cup B) \cup C$. Using a similar approach to the above, we can show that $(A \cup B) \cup C \subset A \cup (B \cup C)$ and then by Theorem 6-2, we have $A \cup (B \cup C) = (A \cup B) \cup C$.

To prove Property *e* of the theorem, we first assume $A \subset B$ and prove that this implies $A \cup B = B$.

- Take any $x \in A \cup B$, then $x \in A\ or\ x \in B$. In the former case, we have $x \in A \subset B \Rightarrow x \in B$. In the latter case, we already have $x \in B$. So, either way we have that $x \in B$. Thus, $A \cup B \subset B$.
- Take any $x \in B$, then x is (by definition) in A or B, i.e., $x \in A \cup B \Rightarrow B \subset A \cup B$.

Since we have shown that $A \cup B \subset B\ and\ B \subset A \cup B$, then $A \cup B = B$.

Going in the other direction, we need to show that $A \cup B = B$ implies that $A \subset B$. So, take any $x \in A$, it then follows that $x \in A \cup B$, but we are given $A \cup B = B$ and thus $x \in B$. So, every element in A is an element of B, i.e., $A \subset B$■

[**Author's Remark**: The above proofs may seem a bit laborious but they are meant to serve as practice for structured logical reasoning.]

Similar properties hold for intersection, as stated in the following theorem.

Theorem 6-4 The following properties hold true concerning set intersection:

a. $A \cap \phi = \phi$
b. $A \cap B = B \cap A \ (commutative)$
c. $A \cap (B \cap C) = (A \cap B) \cap C \ (associative)$
d. $A \cap A = A \ (idempotent)$
e. $A \subset B \ if \ and \ only \ if \ A \cap B = A.$

Proof: The proof technique of these properties is similar to that of Theorem 6-3, i.e., show the set on the left-hand side is a subset of the set on the right-hand side, and vice versa■

There are also set properties that involve both union and intersection. These properties are known as distributive laws.

Theorem 6-5 Distributive Laws for Sets:

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Proof: The proofs are very similar for both properties. So, only a proof of the first law is given. In what follows, each line implies the next (the "implies that" statements are omitted).

We first show $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$:

- Take any $x \in A \cap (B \cup C)$
- $x \in A$ and $x \in B \cup C$
- $x \in A$, and $x \in B \ or \ x \in C$
    - Placement of the comma is important here and in the following line. For example, if the comma is removed, one might interpret that statement as ($x \in A$ and $x \in B$) $or \ x \in C$ which would not be correct.
- $x \in A$ and $x \in B$, or $x \in A$ and $x \in C$
- $x \in (A \cap B) \cup (A \cap C)$

Going the other way, i.e., prove $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$:

- Take any $x \in (A \cap B) \cup (A \cap C)$
- $x \in A \cap B$ or $x \in A \cap C$
- $x \in A$ and $x \in B$, or $x \in A$ and $x \in C$. In either case, $x \in A$.
    - $x \in A$ and $x \in B \Rightarrow x \in A \cap (B \cup C)$

- $x \in A$ and $x \in C \Rightarrow x \in A \cap (B \cup C)$■

For those more visually inclined, Figure 3 may help to see the validity of the first distributed law. As one can see,

- B and C can be united and then intersected with A, or
- A can be individually intersected with B and C, and then take the union of $A \cap B$ *and* $A \cap C$.

The same result (gray area in the figure) is reached in either case.

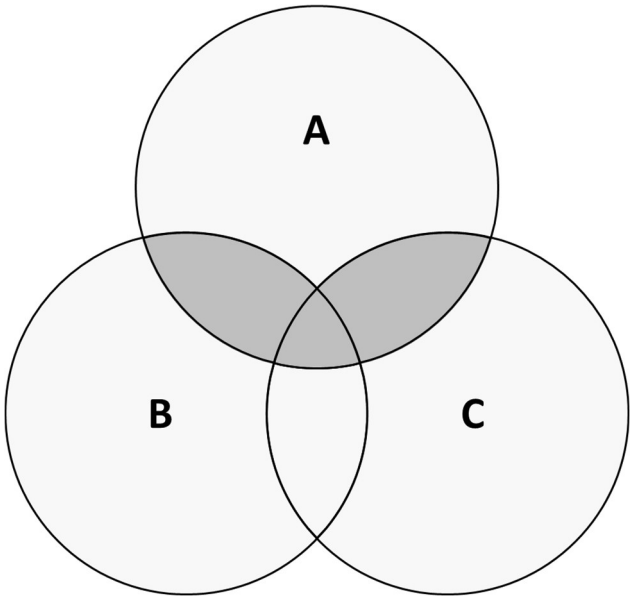


Figure 3. Venn diagram of A ∩ (B∪C) = (A∩B) ∪ (A∩C)

De Morgan's laws entail union, intersection and complements, and are used quite often when dealing with sets:

- $\neg(A \cup B) = \neg A \cap \neg B$
- $\neg(A \cap B) = \neg A \cup \neg B$.

As shown in the following theorem, De Morgan's laws can be generalized to more than 2 sets.

Theorem 6-6 The following properties hold true for $n \geq 1$, where n is an integer

- $\neg(A_1 \cup A_2 \cup \ldots \cup A_n) = \neg A_1 \cap \neg A_2 \cap \ldots \cap \neg A_n$
- $\neg(A_1 \cap A_2 \cap \ldots \cap A_n) = \neg A_1 \cup \neg A_2 \cup \ldots \cup \neg A_n$.

Proof: We only prove one of the two properties. Again, we use the technique of showing any element of the left-side of the equation is also an element of the right-side, and vice versa.

$x \in \neg(A_1 \cup A_2 \cup \ldots \cup A_n) \Leftrightarrow x$ is not in $A_1, A_2, \ldots$ or $A_n \Leftrightarrow x \notin A_1, x \notin A_2, \ldots,$ and $x \notin A_n$

$\Leftrightarrow x \in \neg A_1, x \in \neg A_2, \ldots,$ and $x \in \neg A_n \Leftrightarrow x \in \neg A_1 \cap \neg A_2 \cap \ldots \cap \neg A_n$■

Compared to the proof of Theorem 6-5, we took a bit of a shortcut. Each step implies the next as well as the previous. So, the sequence of logic flows in two directions.

De Morgan's laws also hold true for an infinite number of sets. As an example, consider the infinite collection of sets $A_n = \left[-\frac{1}{n}, 1+\frac{1}{n}\right]$, i.e., the closed interval of real numbers from $-\frac{1}{n}$ to $1+\frac{1}{n}$. Note that as n increases, A_n gets closer and closer to [0, 1], and $\forall n\ [0,1] \subset A_n$.

For those not familiar with the open and closed interval notation:

- $[0,1]$, referred to as a closed interval, includes 0 and 1
- $(0,1)$, referred to as an open interval, does not include 0 or 1.

Applying De Morgan's law, we have

- $\neg(A_1 \cup A_2 \cup A_3 \cup \ldots) = \neg[-1,2] = (-\infty,-1) \cup (2,\infty)$. Keeping in mind that the largest set in the collection is $A_1 = [-1,2]$ and all the other sets are subsets of A_1. So, the union is $[-1,2]$ and the complement is $(-\infty,-1) \cup (2,\infty)$.
- On the other hand, $\neg A_1 \cap \neg A_2 \cap \neg A_3 \cap \ldots$ gives the same result. First, note that $\neg A_n = \left(-\infty, -\frac{1}{n}\right) \cup \left(1+\frac{1}{n}, \infty\right)$. In this case, each $\neg A_{n+1}$ is larger (includes more of the real numbers) than $\neg A_n$. So, the intersection is the of all the $\neg A_n$ sets is the smallest among them, i.e., $\neg A_1 = (-\infty,-1) \cup (2,\infty)$.

There are many more properties of sets which are not covered here. For a more comprehensive discussion of basic set theory, see the book by Halmos [17].

6.5 Equivalence Relationships and Partitions

A **binary relation** between sets A and B is a set R of ordered pairs (a, b), where $a \in A$ and $b \in B$. An element a is related to an element b, if and only if $(a,b) \in R$ (this can also be written as aRb). In terms of notation, the set of all ordered pairs (a,b) where $a \in A$ and $b \in B$ is called the cross product of A and B and is written as $A \times B$. This is basically a mapping between elements of set A and elements of set B.

For example, let $A = B = \mathbb{Z}$ (set of all integers) and let relation S be defined as $\{(a,b): |a-b| = 7\}$ where $|x|$ is the absolute value function (basically, turns negative numbers into positive numbers, and makes no change to positive numbers). S is the set of all pairs of integers that differ by 7, e.g., $(0,7), (-2,5), (-13,-6)$.

A binary relation E on a set X is an **equivalence relation** if and only if the following properties hold true $\forall a,b,c \in X$:

- aEa (Reflexivity)
- aEb if and only if bEa (Symmetry)
- if aEb and bEc then aEc (Transitivity).

Notice that relative to the definition of a binary relation, an equivalence relation assumes A and B are the same set, i.e., X.

Is the relation S, defined above, an equivalence relation? To answer the question, each of the required properties needs to hold true, but we fail on the very first one, i.e., a is not related to itself since $|a-a| = 0$ and not 7.

The following binary relations are equivalence relations:

- "Is equal to" over the set of real numbers.
- "Has the same birthday as" over the set of all people.
- "Is congruent to" over the set of all triangles.
- "$x - y$ is exactly divisible by 2" over the set of integers.

Let E be an equivalence relation on a set A and let $a \in A$. The **equivalence class** of $\boldsymbol{a}$ ***modulo*** $\boldsymbol{E}$ is defined as $[\![a]\!]_E = \{x \in X : xEa\}$.

One interesting thing about an equivalence relation on a set X is that it divides X into a collection of disjoint subsets. It is also true that if set X is divided into a collection of disjoint subsets, it is possible to work backwards and define an equivalence relation that generates the partition of X consisting of the disjoint subsets. These statements are formalized in the following. The proof of Theorem 6-7 is not provided here but can be found in the book by Hrbacek and Jech [18].

Theorem 6-7 For $a, b \in A$ and equivalence relationship E,

a. *aEb (i.e., a is equivalent to b) if and only if $[\![a]\!]_E = [\![b]\!]_E$*
b. *a is not equivalent to b if and only if $[\![a]\!]_E \cap [\![b]\!]_E = \phi$.*

So, an equivalence relationship divides a set into disjoint subsets where all the elements in a given subset are equivalent to each other.

A collection of nonempty sets X is called a **partition** of a set A if the sets in X are pairwise disjoint, and the union of the sets in X equal A. The sets in a partition are called **subdivisions**. This definition does not say anything about equivalence relationships. However, the following theorem relates partitions and equivalence relationships.

Theorem 6-8 If E is an equivalence relationship defined on a set A, the set of equivalence classes on A (denoted A/E) is a partition of A.

Proof: This follows directly from Theorem 6-7.

We can also go in the other direction, i.e., show that a partition defines an equivalence relationship.

Assume X is a partition of a set A. We define the binary relation $E(X, A)$ as follows:

- $(a, b) \in E(X, A)$ if there exist a set $Y \in X$ such that $a \in Y$ and $b \in Y$.

In words, $E(X, A)$ is a binary relation on set A such that a and b are related if and only if they are in the same set Y where Y is one of the subdivisions in X. In fact, $E(X, A)$ is an equivalence relation on A as is shown in the following theorem.

Theorem 6-9 If X is a partition of a set A, then E(X,A) defines an equivalence relation on A.

Proof:

(Reflexivity) Take any $a \in A$. Since A is the union of all the elements in the partition X, there must exists some $Y \in X$ such that $a \in Y$ and thus, $(a, a) \in E(X, A)$.

(Symmetry) If $(a, b) \in E(X, A)$, then there is a $Y \in X$ such that $a \in Y$ and $b \in Y$ which means that, by definition of $E(X, A)$, $(b, a) \in E(X, A)$.

(Transitivity) If $(a, b) \in E(X, A)$ and $(b, c) \in E(X, A)$, then there is Y such that $a \in Y$ and $b \in Y$, and a Z such that $b \in Z$ and $c \in Z$. But the elements of X are disjoint, and we have that $b \in Y$ and $b \in Z$. Thus, it must be that $Y = Z$. So, we have $a \in Y = Z$ and $c \in Y = Z$, which implies $(a, c) \in E(X, A)$ ■

As an example, consider the binary relation defined as follows: for a fixed positive integer n, a is related to b modulo n if $a - b = kn$ for some integer k. For example, if we take $n = 11$, then 3 is related to all of the numbers in the set $\{\dots, -30, -19, -8, 3, 14, 25, 36, \dots\}$. The notation "modulo n" goes back to the book Disquisitiones Arithmeticae (1798) by the famous mathematician Karl Friedrich Gauss (1777-1855). Gauss used the symbol ≡ for the "modulo n" equivalence. So, if a is related to b modulo n, we would write $a \equiv b \ (mod\ n)$. This is an equivalence relationship, as is shown below. For fixed integer n,

- (reflexivity) $a \equiv a \ (mod\ n)$, since $a - a = 0 \cdot n$
- (symmetry) if $a \equiv b \ (mod\ n)$, then there exists k such that $a - b = kn$ but this can be rewritten as $b - a = (-k)n$ and so $b \equiv a \ (mod\ n)$
- (transitivity) if $a \equiv b \ (mod\ n)$ and $b \equiv c \ (mod\ n)$, then there exists k such that $a - b = kn$ and l such that $b - c = hn$. If we solve for b in the first equation, i.e., $b = a - kn$ and substitute into the second equation, we get $a - kn - c = hn$ which can be rewritten as $a = c + (k + h)n$. Thus, $a \equiv c \ (mod\ n)$.

For $n = 11$, what are the equivalence classes for the "mod 11" equivalence relation? There are eleven classes which we can represent as $[\![k]\!] = \{\dots, k - 22, k - 11, k, k + 11, k + 22, \dots\}$, for $k = 0,1,2, \dots, 10$. For example, $[\![3]\!] = \{\dots, -19, -8, 3, 14, 25, \dots\}$.

There are many basic properties that can be proven for modulo arithmetic, some of which are summarized in the following theorem.

Theorem 6-10 Given a fixed integer $n > 1$, and arbitrary integers a, b, c and d, the following holds true:

a. $a \equiv b(mod\ n)$ *and* $c \equiv d(mod\ n) \Rightarrow (a + c) \equiv (b + d)(mod\ n)$ *and* $ac \equiv bd(mod\ n)$
b. $a \equiv b(mod\ n) \Rightarrow (a + c) \equiv (b + c)(mod\ n)$ *and* $ac \equiv bc(mod\ n)$
c. $a \equiv b(mod\ n) \Rightarrow a^k \equiv b^k(mod\ n)$ *for any positive integer k.*

Another example, known as Equivalence Class Partitioning (ECP), comes from the world of software testing. ECP is a software testing technique that divides the input data for a unit software test into partitions of equivalent data from which test cases are derived. The idea is for the test cases (referred to as test vectors) to cover all aspects of the software with no or minimal overlap. The test vectors are typically designed to reveal particular classes of errors. The goal of this approach is to systematically test all aspects of a given software unit, while minimizing the number of test cases.

For a given software unit, a given input vector (data) will test some subset of the instructions. This defines an equivalence relationship (call it T) between input test vectors, i.e., test vectors a and b are equivalent (written as aTb) if and only if a and b test (exercise) the same subset of instructions

for the given software unit. This equivalence relationship partitions the space of test vectors into multiple equivalence classes. For a more complete description of this approach, see the Wikipedia article on Equivalence partitioning [19].

6.6 Paradoxes and Interesting Facts

6.6.1 Hilbert's Paradox of the Grand Hotel

David Hilbert's paradox of the grand hotel (sometimes known as the infinite hotel paradox) illustrates the issues in defining cardinality for infinite sets. [David Hilbert (1862 – 1943) is considered one of the most influential mathematicians of the 19th and early 20th centuries. Hilbert discovered and developed a broad range of fundamental ideas in mathematics.]

The story goes something like this: a very efficient manager runs an infinite hotel with all the rooms occupied. The rooms are numbered 1, 2, 3, …

A bus arrives with a finite number of guests (say 10) who are looking for rooms. The hotel manager moves the existing hotel guests to different rooms as follows:

Table 15. Reassignment to make for 10 additional rooms

Before Reassignment	1	2	3	4	5	6	7	…
After Reassignment	11	12	13	14	15	16	17	…

It should be emphasized:

- Every existing customer has a room after the reassignment. In general, the guest in room n is now in room $n + 10$.
- No room has more than one guest assigned to it.
- Rooms #1-10 are now free for assignment to the new guests.

So, countable infinity plus 10 (or any finite number) is still countable infinity.

Next, a train with a countably infinite number of people arrives at the hotel. Again, the hotel manager is able to accommodate the new guests by reassigning the existing guests as follows:

Table 16. Reassignment to make for a countably infinite number of additional rooms

Before Reassignment	1	2	3	4	5	6	7	…
After Reassignment	2	4	6	8	10	12	14	…

In this reassignment, each existing guest is assigned to a unique room. The guest in room n is moved to room $2n$. This leaves the odd numbered rooms (an infinite number) available for the new guests who just arrived on the train. If we assume the guests on the train are numbered $1, 2, 3, \ldots$ then the guest assigned number r on train is put in room $2r - 1$ in the hotel.

So, if we have two countably infinite sets, the union is still countably infinite.

We could next try to accommodate the passengers from a finite number of trains each with a countably infinite number of passengers who are looking for rooms but let's go to the more complex case of a countably infinite number of trains, each with a countably infinite number of

passengers. Our clever hotel manager is still able to accommodate all the guests via arrangement of the existing hotel guests. One approach is as follows:

Table 17. Countably infinite number of trains – Arrangement #1

New room assignment for existing guests	2^1	2^2	2^3	2^4	2^5	…
Room assignment for passengers in Train #1	3^1	3^2	3^3	3^4	3^5	…
Room assignment for passengers in Train #2	5^1	5^2	5^3	5^4	5^5	…
Room assignment for passengers in Train #3	7^1	7^2	7^3	7^4	7^5	…
…						

The mapping in Table 17 makes use of prime numbers, i.e., a positive integer that is only divisible by 1 and itself, and the fact that there are an infinite number of primes (as we will prove in Theorem 9-21).

Concerning the mapping in Table 17, the nth passenger in the mth train is placed in Room p_{m+1}^n, i.e., the room number is the m+1st prime number raised to the nth power. This scheme leaves many empty rooms, basically any room that is a multiple of two or more primes (raised to a power), e.g., 6^2 x 7^3 = 12348 is empty after the rearrangement of existing guests and placement of new guests.

Other arrangements are possible. For example, define the set $A = \{ (x, y): x \in \mathbb{N}, y \in \mathbb{N}\}$, i.e., the set of possible pairs of natural numbers (positive integers including 0). A is also a countable set and can be put into a one-to-one correspondence with the natural numbers (see Theorem 6-11). So, we can label the elements of A as $a_1, a_{2,} a_3, \ldots$

If $a_n = (x, y)$, we assign the yth guest on the xth train to the nth room. Existing guests are considered to be on Train 0. In this scheme, each guest is assigned a unique room. Further, for each room there is a guest since the mapping between the set A and the natural numbers is a one-to-one correspondence.

So, we now see that a countably infinite set of sets (each having a countably infinite number of elements) is also countably infinite. The reader may think at this point that "countably infinite" (i.e., one-to-one correspondence with the natural numbers) is the only type of infinite. As we shall see in the following subsections, this is far from the case.

Theorem 6-11 The set $\{ (x, y): x \in \mathbb{N}, y \in \mathbb{N}\}$ is countable (i.e., countably infinite).

Proof: We need to show a one-to-one correspondence with $\mathbb{N}$. The following table lists all elements of A in a systematic manner.

Table 18. Listing of all pairs of natural numbers

(0,0)	(0,1)	(0,2)	(0,3)	(0,4)	(0,5)	…
(1,0)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	…
(2,0)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	…
(3,0)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	…
(4,0)	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	…
…	…	…	…	…	…	

The dashed line shows an ordering (listing) of the elements of A. This listing demonstrates that A is countable■

Just for the record, the set of all fractions (rational numbers) is also countably infinite and as such can be put into a 1-1 correspondence with the natural numbers. The proof is very similar to that of Theorem 6-11. We just write each fraction $\frac{a}{b}$ as a pair (a, b). The only issues in using the diagonalization approach are as follows:

- Division by 0, e.g., in pairs of the form $(a, 0)$. So, just remove the first column from Table 18.
- There are also duplicates, e.g., $(1,2)$ and $(2,4)$ both represent $\frac{1}{2}$. In this case, we keep the representation in the most reduced form and delete all the duplicates.

6.6.2 Cantor's Diagonalization Argument

What is now known as Cantor's diagonalization argument was published in 1891 by Georg Cantor as an existence proof of a set whose cardinality is larger than that of the natural numbers. In other words, Cantor defined a set that cannot be put into a one-to-one correspondence with the natural numbers.

Cantor's construction is fairly straightforward. First, take the set of all possible binary sequences (call it B). An example element of B is $(1, 0, 1, 0, 1, 0, \dots)$.

Assume that B is countable. This assumption means there is an enumeration of B which we write as $b_1, b_2, b_3, b_4, \dots$

We can list the elements of B as follows:

$$b_1 = (1, 1, 1, 0, 1, 1, \dots)$$
$$b_2 = (1, 1, 1, 0\ ,0\ , 0, \dots)$$
$$b_3 = (0, 0, 1\ ,1\ 0, 0, \dots)$$
$$b_4 = (1, 0, 0, 0, 1, 1, \dots)$$
$$\dots$$

This is just an example of the possible enumerate of the elements in B. Any enumeration will suffice to make the point here.

Now, the trick is to define another sequence $x = (x_1, x_2, x_3, \dots)$ such that its x_i is the opposite from that of b_i. So, if $b_i = 1$ then $x_i = 0$, and if $b_i = 0$ then $x_i = 1$. For the example above, $x = (0, 0, 0, 1, \dots)$. Basically, one goes down the diagonal of the list below and flips the value of each of the digits in bold to form a new sequence x.

$$b_1 = (\mathbf{1}, 1, 1, 0\ 1, 1, \dots)$$
$$b_2 = (1, \mathbf{1}, 1, 0\ ,0\ , 0, \dots)$$
$$b_3 = (0, 0, \mathbf{1}\ ,1\ 0, 0, \dots)$$
$$b_4 = (1, 0, 0, \mathbf{0}, 1, 1, \dots)$$
$$\dots$$

The new sequence x is different from each of the b_n for whatever value of n one chooses, because x (by definition) differs from b_n in (at least) the n^{th} digit. So, we have that x is a binary sequence but $x \notin B$ which is a contraction since B was defined as the set of all binary sequences. Thus, our assumption that B is countable must be incorrect. So, we must conclude that B is an uncountable set.

Using a similar technique to the above, we can prove that the set of real numbers in the closed interval [0, 1] is uncountable. The proof goes as follows:

Assume that [0,1] is countably infinite. This implies there is a 1-1 mapping between the natural numbers $\mathbb{N}$ and [0,1], i.e., we can list (enumerate) real numbers in [0,1]. In decimal notation, the list would look like:

$$r_1 = .\mathbf{3}4820\ \dots$$
$$r_2 = .9\mathbf{3}341\ \dots$$
$$r_3 = .34\mathbf{8}33\ \dots$$
$$r_4 = .776\mathbf{5}3\ \dots$$
$$r_5 = .9973\mathbf{7}\ \dots$$
$$\dots$$

The above is just part of one example list (any listing will do).

Define $x \in [0,1]$ as follows:

- If the nth decimal spot in r_n is less than 9, then define the nth decimal spot of x as $r_n + 1$.
- If the nth decimal spot in r_n is 9, then define the nth decimal spot of x as 0.

For the example listing above, x would be $.44968\dots$ and by construction, different in at least one digit with every element in the list.

By construction, x is different from every number in the list. So, we have a contradiction to the assumption that all the real numbers in the interval [0,1] can be enumerated. Thus [0,1] is uncountable. Further, $\tan\left(\pi x - \frac{\pi}{2}\right)$ maps the interval (0,1) in a 1-1 manner to $(-\infty, \infty)$, i.e., set set real numbers $\mathbb{R}$. Thus $\mathbb{R}$ is also uncountable and of the same cardinality as (0,1) which is the same cardinality as [0,1].

6.6.3 The Cantor Set

The Cantor set (discovered by Henry John Stephen Smith [20] in 1874 and popularized by Georg Cantor [21]) is a subset of the interval [0,1] with some unexpected properties. The set is fairly easy to describe but the analysis is a bit complex.

The Cantor set is defined by an iterative process. At each iteration additional points are removed.

- At the first step, the entire interval [0,1] is included.
- At the second step, the interval $\left(\frac{1}{3},\frac{2}{3}\right)$ is removed.
- At the third step, the intervals $\left(\frac{1}{9},\frac{2}{9}\right)$ and $\left(\frac{7}{9},\frac{8}{9}\right)$ are removed.
- At the fourth step, the interval $\left(\frac{1}{27},\frac{2}{27}\right), \left(\frac{7}{27},\frac{8}{27}\right), \left(\frac{19}{27},\frac{20}{27}\right)$ and $\left(\frac{25}{27},\frac{26}{27}\right)$ are removed.

The process continues indefinitely and what remains is defined to be the Cantor set. The basic idea is to remove the middle third interval from each remaining interval at each step in the process. The first four iterations are shown in Figure 4.

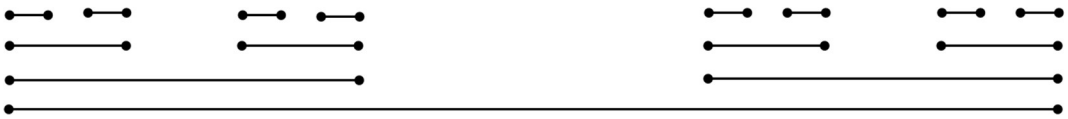


Figure 4. First four iterations of the Cantor Set

Is there anything left when the process is completed? One approach is to add the length of the removed intervals. However, to do this, we first need to prove a basic formula for the sum of a geometric series.

Theorem 6-12 The sum of the geometric series $1 + r + r^2 + r^3 + r^4 + \cdots$ is $\frac{1}{1-r}$ where $0 < r < 1$.

Proof: First, determine a formula for the sum of the first n terms.

Let $s = 1 + r + r^2 + r^3 + \cdots + r^{n-1}$, then it follows that $rs = r + r^2 + r^3 + \cdots + r^n$.
Subtracting the equation for rs from s gives $s - rs = s(1-r) = 1 - r^n$ which implies $s = \frac{1-r^n}{1-r}$.

Since $0 < r < 1, r^n$ approaches 0 as n approaches infinity and so we have the desired result that

$1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r}$ ■

Regarding the Cantor set, lengths of the intervals removed at each step are $\frac{1}{3},\frac{2}{3^2},\frac{2^2}{3^3},\frac{2^3}{3^4},\ldots$

Taking the sum and making use of Theorem 6-12:

$$\frac{1}{3}+\frac{2}{3^2}+\frac{2^2}{3^3}+\frac{2^3}{3^4}+\cdots=\frac{1}{3}\left(1+\frac{2}{3}+\left(\frac{2}{3}\right)^2+\left(\frac{2}{3}\right)^3+\cdots\right)=\frac{1}{3}\left(\frac{1}{1-\frac{2}{3}}\right)=\frac{1}{3}.3=1$$

The length (or "measure" to be more precise) of the Cantor set is 0 since the measure of the removed intervals is 1 and of course, the measure of [0,1] is 1. This means that the Cantor set does

not contain any interval of non-zero length. However, the Cantor set is not empty! In fact, the endpoints for all the intervals at each level are members of the Cantor set. There are also numbers, other than the endpoints of removed intervals, that are in the Cantor set. For example, it can be shown that $\frac{1}{4}$ is in the Cantor set. For a proof of this fact, see the article by Belcastro and Green [22].

In order to fully describe all the numbers in the Cantor set, it is helpful to work in base 3 (for reasons that will become clear). As a quick review, recall that in base 3 there are three one-digit numbers, e.g., 0,1 and 2. All numbers are written as positive or negative powers of these three digits. In what follows, numbers written in base 3 will have a subscript of 3 (e.g., 12102_3) and numbers base 10 will be written without any subscript. Some examples,

- $120_3 = 1 \cdot 3^2 + 2 \cdot 3 + 0 \cdot 1 = 9 + 6 + 0 = 15$
- $.212_3 = 2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3^2} + 2 \cdot \frac{1}{3^3} = \frac{2}{3} + \frac{1}{9} + \frac{2}{27} = \frac{23}{27}$.

As it turns out, the Cantor set contains all numbers in the interval [0,1] whose base 3 representation consists of zeros and twos, with no ones. To see this, we consider the representation of the Cantor set elements at each iteration.

Now consider the representation (base 3) of the numbers remaining after the first $(\frac{1}{3}, \frac{2}{3})$ is removed.

- All the numbers in $[0, \frac{1}{3}]$ can be written in the form .0xyz…, noting that $\frac{1}{3}$ can be written as .0222… or as .1 (we use the former representation).
- All the numbers in $[\frac{2}{3}, 1]$ can be written in the form .2xyz… since each number in this interval requires a $\frac{2}{3}$ term and some smaller powers of $\frac{1}{3}$ (represented as "xyz…" in the above). Note that $\frac{2}{3}$ can be written as .2 or as .1222… (we use the former representation). Also, 1 can be written in base 3 as 1 or .222… (we use the latter).

At the next step, when $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ are removed, the most significant digit (in base 3) is not affected, i.e., all the numbers in $[0, \frac{1}{9}]$ and $[\frac{2}{9}, \frac{1}{3}]$ still have a 0 as their first digit and all the numbers in $[\frac{2}{3}, \frac{7}{9}]$ and $[\frac{8}{9}, 1]$ still have a 2 as their first digit.

- All the numbers in $[0,\frac{1}{9}]$ have 0 as their second digit since each number in this interval does not have a $\frac{1}{3^2}$ term (they are all smaller than $\frac{1}{9}$). So, all numbers in $[0, \frac{1}{9}]$ are of the form .01xyz … Also, we have the same double representation issue with $\frac{1}{9}$, i.e., .01 or .00222… (we use the latter).
- All the numbers in $[\frac{8}{9}, 1]$ have 2 as their second digit since each number in this interval can be written as $\frac{8}{9} = \frac{2}{3} + \frac{2}{9}$ plus smaller order terms.
- Similar arguments can be made for the numbers in $[\frac{2}{9}, \frac{1}{3}]$ and $[\frac{2}{3}, \frac{7}{9}]$.

Continuing in this manner, it is evident that all the numbers in the Cantor set can be written in base 3 using only the digits 0 and 2, or perhaps more to the point, the only numbers removed from [0, 1] when creating the Cantor set are those whose representation contain some ones in their base 3

representation. That is to say that the Cantor set is the set of all numbers in [0,1] whose base 3 representation contains only twos and zeros.

Further, it is possible to map each element in the Cantor set to a binary number simply by changing each appearance of a 2 to a 1. For example, $.2202_3$ is mapped to $.1101_2$. However, there are cases where two elements of the Cantor set are mapped to the same binary number, e.g.,

- $\frac{1}{3}$ = .022 …. base 3 gets mapped to .011… = .1 base 2
- $\frac{2}{3}$ = .2 base 3 gets mapped to .1 base 2.

In general, this double mapping is true for all points in the Cantor set that are at opposite ends of a removed open interval. The mapping is “onto” (for every binary number in [0,1] there is at least one ternary (base 3) number in the Cantor set that gets mapped to it) but not one-one (as we saw in the previous example). So, the cardinality of the Cantor set is at least that of the interval [0, 1]. On the other hand, the Cantor set was created as a subset of [0 ,1] and so the cardinality of the Cantor set cannot be larger than that of the interval [0,1]. Thus, the cardinality is equal to that of [0, 1], i.e., uncountable.

So, the Cantor set is of measure 0 and yet uncountable with the same cardinality as [0, 1].

[**Author’s Remark**: If for two infinite sets A and B, there exists a mapping from A to B such that for every element of B this is a unique element of A that is mapped to it, and a similar function from B to A, then A and B are of the same cardinality. This result is known as the Schröder–Bernstein theorem [23]. Concepts such as onto and 1-1 functions are covered further in Section 7 on functions.]

6.6.4 Russell's Paradox

As discussed earlier in this document, a set can contain other sets. For example, consider B = { 1, 2, 3,} and A = { B, x, y, z }. A is a valid set. It is possible to push this even further and define a set that contains itself, e.g., C = { C, a, b, c}. C is also a valid set. This may lead one to ask, “what is the formal definition of a set?” **[Author’s Remark**: Read on if you are interested in the answer, but you may be sorry as the answer is complex.]

In his book Naïve Set Theory[17] , Paul Halmos states

> One thing that the development will not include is a definition of sets. The situation is analogous to the familiar axiomatic approach to elementary geometry. That approach does not offer a definition of points and lines, but rather, it describes what it is that one can do with those objects. The semi-axiomatic point of view adopted here assumes that the reader has the ordinary, human, intuitive (and frequently erroneous) understanding of what sets are; the purpose of the exposition is to delineate some of the many things that one can correctly do with them.

As translated from German in the Wikipedia article on Sets [24], Georg Cantor provided the following definition of a set [25]:

> A set is a gathering together into a whole of definite, distinct objects of our perception [Anschauung] or of our thought—which are called elements of the set.

Bertrand Russell (philosopher, logician, mathematician, historian, writer, essayist, social critic, political activist, and Nobel laureate) proposed the following set as a counterexample to Cantor's definition of set:

> Let R be the set of all sets which do not contain themselves.

R can be written more formally as $R = \{ A \mid A \notin A\}$. Is R an element of R?

- If R is an element of R, then it contains itself but this violates the definition of R which only contains sets that do not contain themself.
- If R is not an element of R, then by definition, R must be in R (a contradiction).
- In terms of notation, $R \in R \Leftrightarrow R \notin R$. So,

R is a defined collection of things (fits the definition of "set" by Cantor) but this leads to a contraction, hence the paradox.

At the time this paradox was introduced, mathematicians agreed that there was a problem with the definition of "set" proposed by Cantor. Several solutions to the problem were proposed. One solution entailed an axiomatic formulation of set theory which is free of paradoxes (including the one posed by Russell). This formulation is known as Zermelo–Fraenkel Set Theory [26]. Zermelo–Fraenkel set theory is comprised of the axioms stated below. These are stated as informally as possible. From these relatively few axioms (basically statements that are given as true), one can formally develop a theory of sets.

- Axiom of extensionality – Two sets are equal if they have the same elements.
- Axiom schema of specification – Every subclass of a set that is defined by a predicate is itself a set.
    - "Class" is not formally defined in Zermelo–Fraenkel set theory. Informally, a class is a collection of sets that can be unambiguously defined by a property that all its members share. Every set is a class, but there are classes (called proper classes) that are not sets, e.g., the class of all sets or the class of all sets which do not contain themselves (from Russell's paradox).
    - A predicate is a statement or assertion about the elements in a set. For the set of natural numbers $\mathbb{N}$, one might impose the predicate "numbers evenly divisible by 7." The induced subset of $\mathbb{N}$ under this predicate is { 0, 7, 14, 21, …}.
- Axiom of regularity – Every non-empty set A contains a member x such that $A \cap x = \phi$, i.e., A and x are **disjoint**.
    - If A has an element x that is not a set, then A and x cannot have any elements in common since x (not being a set) has no elements. Thus, in this case, $A \cap x = \phi$.
    - On the other hand, there are cases where all the elements of A are sets. For example, take $A = \{B, C\}, B = \{B, x\}$ and $C = \{C, y\}$ then $A \cap B = B$ and $A \cap C = C$. Thus, there does not exist an element of A such that its intersection with A is empty. So, A violates the axiom of regularity.
- Axiom of pairing – Given any two sets (A and B), there is a set whose elements are exactly the two given sets.

- Axiom of union – For any set A, there is a set which consists of just the **elements of the elements** of that set A.
    - For example, if $A = \{\{a, b, c\}, \{c, d, e\}, \{e, f\ g\}\}$ then the set consisting solely of elements of elements of A is $\{a, b, c, d, e, f, g\}$.
- Axiom schema of replacement – For a function f and any set A, $f(A)$ is also a set.
    - Functions are defined in Section 7 of this book.
- Axiom of power set – Given any set A, there is a set that consists of all subsets of A. This set is known as the power set of A.
    - For example, if A = { 1, 2, 3}, the power set of A is $\{\phi, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$.
    - If a set A has n elements, then the power set of A has 2^n elements.
- Axiom of infinity – There exists a set with infinitely many elements.
    - This set can be constructed by taking progressive power sets of the empty set, i.e., $\{\ \phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}, \ldots\}$. The point is that once the existence of an empty set is agreed, an infinite set can be constructed.
- Axiom of choice (informal definition) – Given any collection of sets, with each set containing at least one element, it is possible to select exactly one element from each set, even if the collection is infinite.
    - This may seem obvious and trivial, but it is not. For example, the monograph by H. Rubin and J.E. Rubin [26] lists over 250 propositions which are equivalent to the axiom choice.

[**Author's Remark**: Don't be alarmed by the complexity of the above set of axioms. We will only need a few of these to address Russell's paradox and one other theorem. The point in listing the full set of axioms is to illustrate that a large amount of mathematics (in the case of Set Theory) can be based on a relatively small number of assumptions.]

Back to Russell's paradox …

As Patrick Suppes states in Section 1.3 of his book *Axiomatic Set Theory* [28], Cantor did not proceed from an explicit collection of axioms when doing his work on set theory. However, Cantor implicitly relied on the following three axioms for many of his proofs:

- Axiom of extensionality (as defined above)
- Axiom of abstraction – For a given property (call it p), there exists a set Y whose elements are exactly those that have the given property. This can be written in notation as $(\exists Y)(\forall x)\big(x \in Y \Leftrightarrow p(x)\big)$, where ∃ means "there exists", ∀ means "for every" and $p(x)$ means the property p holds true for x.
- Axiom of choice (as defined above).

As Suppes describes in his book, it is the axiom of abstraction that caused the problem exposed by Russell's paradox. (Do not confuse the axiom of abstraction with the axiom of specification from the Zermelo–Fraenkel axioms. The two axioms are not the same.)

In 1901, Bertrand Russell discovered that a contradiction could be derived from the axiom of abstraction by considering the set of all things which have the property of not being members of themselves. Let $p(x)$ be the proposition "x is not a member of itself" which can be written as $p(x) = \neg(x \in x)$. Making the substitution for $p(x)$ in formula for the axiom of abstraction, we get $(\exists Y)(\forall x)\big(x \in Y \Leftrightarrow \neg(x \in x)\big)$. There is no restriction on x and we can set it to anything. Let $x = Y$ to get $\big(Y \in Y \Leftrightarrow \neg(Y \in Y)\big)$ which is essentially the same contradiction as found in Russell's paradox. So, while on the surface it seems reasonable to assume that for a given property there exists a set of entities having that property (i.e., the axiom of abstraction), Russell's paradox shows this assumption is not valid.

The problem is solved by the axiom schema of specification. The formal notation for axiom schema of specification looks very similar to that of the axiom of abstraction, i.e.,

$$(\exists Y)(\forall x)\big((x \in \boldsymbol{Y} \subset \boldsymbol{Z}) \Leftrightarrow (\boldsymbol{x} \in \boldsymbol{Z}) \text{ and } p(x)\big).$$

The differences are highlighted in bold. In the case of the axiom schema of specification, one needs to state set Z as part of the conditions. So, $p(x)$ and Z are given, and the axiom states that Y (subset of Z) exists and satisfies the formula. In this case, x cannot be assigned to anything we want. There is the condition that x must be an element of Z.

So, if as before when generating Russell's paradox, we choose $p(x) = \neg(x \in x)$ and $x = Y$, and substitute into the formula for the axiom schema of specification, we get

$$(\exists Y)\big(Y \in Y \Leftrightarrow (Y \in Z) \text{ and } \neg(Y \in Y)\big).$$

A value of Y that satisfies the above is simply $Y = Z$ which gives

$$\big(Z \in Z \Leftrightarrow (Z \in Z) \text{ and } \neg(Z \in Z)\big)$$

The left-side is false regardless of the choice of Z (see Theorem 6-13) and the right-side is also false (a contradiction). So, we have a false statement implying and being implied by a false statement which by basic logic is a true statement.

Theorem 6-13 In Zermelo–Fraenkel set theory, no set is a subset of itself.

Proof: Take any set A and consider {A}. (Note that A and {A} are not the same set.) By the axiom of pairing, the unique set whose members are A and the empty set is {A}, and thus {A} is a valid set in Zermelo-Fraenkel set theory.

By the axiom of regularity, there must be an element of {A} which is disjoint from {A}, but the only element of {A} is A. Thus, it must be that $A \cap \{A\} = \phi$. However, since $A \in \{A\}$, we cannot have $A \in A$ too (by the definition of disjoint). Since there was no condition on the selection of A (could be any set), we have proven that in general no set can have itself as a member (at least not under the axioms of Zermelo-Fraenkel set theory) ■

6.6.5 Ross–Littlewood Paradox

The Ross–Littlewood paradox (first posed by John E. Littlewood and extended by Sheldon Ross) is a hypothetical problem in set theory and logic designed to illustrate the paradoxical nature of something called "supertasks."

Start with a large empty container and an infinite supply of numbered balls. Proceed as follows:

- Step 1: (1 minute before a given time T) add the first ten balls (numbered 1-10) and then remove Ball #1.
- Step 2: (30 seconds before a given time T) add Balls #11-20 and then remove Ball #2
- Step 3: (15 seconds before a given time T) add Balls #21-30 and then remove Ball #3
- Continue this process indefinitely with each step being done half as close to time T as the previous.

How many balls are in the vase when the task is finished?

Infinite: One school of thought favors an infinite number of balls since more balls are being added than subtracted at each step and there are an infinite number of steps.

Zero: If you think any balls are left in the container at the end of the process, just give the number of one such ball (say N) but then Ball N is removed at the Nth step.

There are several issues with the problem statement. First of all, the container cannot be infinite (there is no such thing). We could try to fix that by saying that the container is an abstract set (call it X) that is represented on a computer and the elements of Set X are just the numbers of the balls. Set X is updated by the computer at each step. But even with this modification to the problem statement, computational and storage limits will be reached:

- The intervals keep getting smaller and smaller, and some processing (albeit very little) is required to update Set X. Eventually, a theoretical computing limit will be reached.
- Even with various shorthand representations, the size of Set X will grow larger than what can be stored on any possible computer.

[Author's Remark: From a practical point of view, it is impossible to execute what is described in the problem statement. The problem does not neatly match any of the paradox categories in Section 5.5. In my view, the closest fit is to a falsidical paradox.]

6.6.6 Zeno's Paradox of Motion

This is not a set theory paradox but it fits here since it is similar to the Ross–Littlewood paradox in that both are supertasks.

Assume someone is to run 100 meters (from Point A to B). Zeno's paradox claims that the run must first cover 50 meters, then another 25 meters, followed by half that (i.e., 12.5 meters) and so on, for an infinite number of tasks. Since this requires an infinite number of tasks, the claim is that the runner never reaches Point B.

Of course, we know this is wrong but what is the precise explanation?

Make the reasonable assumption that the runner travels at the same speed of 5 meters/second. The first task takes 10 seconds, the second task takes 5 seconds, the next takes 2.5 seconds, and so on. This gives the infinite sum $10 + \frac{10}{2} + \frac{10}{2^2} + \frac{10}{2^3} + \cdots = 10\left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots\right) = 10\left(\frac{1}{1-\frac{1}{2}}\right) = 20$, making use of the formula for the sum of a geometric series (Theorem 6-12). Thus, the infinite number of tasks take exactly 20 seconds (which we already knew before the confusion caused by Zeno).

So, what appears to be absurd at first glance can, in fact, be proven to be a false statement. Thus, this is a falsidical paradox.

6.7 Exercises

1. Show that the cardinality of the set of positive even integers {2, 4, 6, …} is the same as the cardinality of the set of all integers (both positive and negative), i.e., {…, -2, -1, 0, 1, 2, …}. ***Hint****: Define a one-to-one mapping between the two sets.*
2. For a finite set A, define $n(A)$ as the number of elements in A. For example, if $A = \{a, b, c, d, e\}$, then $n(A) = 5$. For two finite sets A and B, convince yourself that $n(A \cup B) = n(A) + n(B) - n(A \cap B)$. ***Hint****: Try drawing the associated Venn diagram for an example.*
3. Using the result from Exercise #2 and the distributive law for sets, derive a formula for $n(A \cup B \cup C)$. ***Hint****: From Exercise #2, we have* $n\big((A \cup B) \cup C\big) = n(A \cup B) + n(C) - n\big((A \cup B) \cap C\big)$. *From Theorem 6-5, we have that* $(A \cup B) \cap C) = (A \cap C) \cup (B \cap C)$. *Apply the result from Exercise #2 to both* $n(A \cup B)$ *and* $n\big((A \cap C) \cup (B \cap C)\big)$.
4. In a group of 100 people, exactly 40 people play chess, exactly 30 play tennis and 10 play both chess and tennis. How many people play neither chess nor tennis? ***Hint****: Use the result from Exercise #2 to determine the number of people who either play chess or tennis, and then subtract this number from 100.*
5. Describe in words and write down some elements of the power set of the set $\{1, 2, 3, \ldots, n\}$.
6. Prove that the power set of the set of natural numbers $\mathbb{N}$ is uncountable. ***Hint:*** *Assume the power set of the natural numbers is countable and enumerate the elements of the power set, i.e.,* $A_1, A_2, A_3, \ldots$ *Define the set* $A = \{i \in \mathbb{N} \mid i \notin A_i\}$ *and note that A cannot be in the assumed enumeration of the power set of* $\mathbb{N}$. *So,* $i \notin A$ *if* $i \in A_i$, *and* $i \in A$ *if* $i \notin A_i$. *Thus,* $A \neq A_i$ *for every value of* i. *This is similar to the diagonalization process described earlier in this document.*
7. The difference of A and B is defined as the set $A - B = \{x \in A \mid x \notin B\} = A \cap \neg B$. Prove that $(A \cup B) - (A \cap B) = (A - B) \cup (B - A)$. ***Hint and Answer****: Take any element* x *on one side of the equation and reason that it must be a member of the other side of the equation, and vice versa. Another approach is to use the various set properties, i.e.,* $(A \cup B) - (A \cap B) = (A \cup B) \cap \neg(A \cap B) = (A \cup B) \cap (\neg A \cup \neg B) = (A \cap \neg A) \cup (A \cap \neg B) \cup (B \cap \neg A) \cup (B \cap \neg B) = \phi \cup (A \cap \neg B) \cup (B \cap \neg A) \cup \phi = (A \cap \neg B) \cup (B \cap \neg A) = (A - B) \cup (B - A)$.

7 Boolean Algebra

No doubt the reader has noticed very similar laws for propositional logic and sets. This is no accident. In fact, propositional logic and sets are examples of something called a Boolean algebra.

In what follows, we define Boolean algebras, state and prove some general properties about Boolean algebras and then give examples of Boolean algebras. The key point here is that we only need to state and prove theorems once in the context of Boolean algebras. The theorems then apply to all Boolean algebras of which propositional logic and sets are examples.

7.1 Definitions

A **binary operation** maps two elements of a set to exactly one element in the same set. A **unary operation** maps one element of a set to exactly one other element of the same set.

For example, the conjunction operation from propositional logic is a binary operation and the negation operation is a unary operation.

A **Boolean algebra** is a system which consists of a set $\mathcal{A}$ with two associated binary operations named meet ($\sqcap$) and join ($\sqcup$), a unary operation called completion ($\sim$) and which obeys the following axioms (laws):

- $A \sqcup B = B \sqcup A$ and $A \sqcap B = B \sqcap A$, for all $A, B \in \mathcal{A}$ (Commutative laws)
- $A \sqcup (B \sqcap C) = (A \sqcup B) \sqcap (A \sqcup C)$ and $A \sqcap (B \sqcup C) = (A \sqcap B) \sqcup (A \sqcap C)$, for all $A, B, C \in \mathcal{A}$ (Distributive laws)
- There exist elements in $\mathcal{A}$ (denoted by 0 and 1) such that $A \sqcup 0 = A$ and $A \sqcap 1 = A$, for all $A \in \mathcal{A}$ (Identity laws)
- For all $A \in \mathcal{A}$, $A \sqcup {\sim}A = 1$ and $A \sqcap {\sim}A = 0$ (Completion laws).

From the above axioms, many properties can be proven for Boolean algebras. Only a small subset of such properties is covered in this document. For a more comprehensive discussion see the book by Solomon [30].

From meet, join and completion, other operations can be defined, e.g.,

- $A \Rightarrow B$ is defined as $({\sim}A) \sqcup B$
- $A \uparrow B$ (i.e., NAND) is defined as ${\sim}(A \sqcap B)$.

7.2 Theorems

The following theorem cuts our work in half with regard to proving various properties of Boolean algebras. For any given true statement in Boolean algebra, its dual statement (derived by interchanging $\sqcup$ and $\sqcap$, and 0 and 1) is also a true statement.

Theorem 7-1 (Principle of Duality) Each statement that can be derived from the axioms of Boolean algebra gives rise to another true statement in which $\sqcup$ and $\sqcap$, and 0 and 1 are interchanged.

Note: In the proofs that follow, the description on a given line of proof refers to the law that justifies going from the previous line to the current line. In the proof immediately below, this fact is called out by the phrase “applied to the previous line of the proof”. In subsequent proofs, the same statement applies but is omitted for brevity.

For example, consider the statement $A \sqcup A = A$ which can be derived from the axioms of Boolean algebra as follows:

A

$= A \sqcup 0$ (identity law)

$= A \sqcup (A \sqcap \sim A)$ (completion law applied to previous line of proof)

$= (A \sqcup A) \sqcap (A \sqcup \sim A)$ (distributive law applied to previous line of proof)

$= (A \sqcup A) \sqcap (1)$ (completion law applied to previous line of proof)

$= A \sqcup A$ (definition of 1 applied to previous line of proof).

The duality principle tells us that if we replace $\sqcup$ with $\sqcap$ in the statement that we just proved, we get another true statement, i.e., $A \sqcap A = A$. For future reference, this result is captured in the following theorem.

Theorem 7-2 (Idempotent laws) For a Boolean algebra $\mathcal{A}$ and any $A \in \mathcal{A}$, $A \sqcup A = A$ and $A \sqcap A = A$.

As another example of the utility of the duality principle, consider the following theorem concerning the distributive law for four elements of a Boolean algebra.

Theorem 7-3 (Extended Distributive Laws) For a Boolean algebra $\mathcal{A}$ and any $A, B, C, D \in \mathcal{A}$,

- $(A \sqcup B) \sqcap (C \sqcup D) = (A \sqcap C) \sqcup (B \sqcap C) \sqcup (A \sqcap D) \sqcup (B \sqcap D)$
- $(A \sqcap B) \sqcup (C \sqcap D) = (A \sqcup C) \sqcap (B \sqcup C) \sqcap (A \sqcup D) \sqcap (B \sqcup D)$.

Proof:

$(\mathrm{A} \sqcup \mathrm{B}) \sqcap (\mathrm{C} \sqcup \mathrm{D}) = [(\mathrm{A} \sqcup \mathrm{B}) \sqcap \mathrm{C}] \sqcup [(\mathrm{A} \sqcup \mathrm{B}) \sqcap \mathrm{D}]$ (distributive law)

$= (\mathrm{A} \sqcap \mathrm{C}) \sqcup (\mathrm{B} \sqcap \mathrm{C}) \sqcup (\mathrm{A} \sqcap \mathrm{D}) \sqcup (\mathrm{B} \sqcap \mathrm{D})$ (distributive law).

From the duality principle, we also get $(\mathrm{A} \sqcap \mathrm{B}) \sqcup (\mathrm{C} \sqcap \mathrm{D}) = (\mathrm{A} \sqcup \mathrm{C}) \sqcap (\mathrm{B} \sqcup \mathrm{C}) \sqcap (\mathrm{A} \sqcup \mathrm{D}) \sqcap (\mathrm{B} \sqcup \mathrm{D})$ ■

In what follows, the term “extended distributive law” is used to refer to Theorem 7-3 and its dual.

The astute reader may have noticed that the associative laws are missing from the definition of a Boolean algebra. However, the associative laws can be derived from the other axioms. To prove this, first we need to prove several results that will be used in the proof of the associative laws.

Theorem 7-4 (Domination laws) For Boolean algebra $\mathcal{A}$ and any $A \in \mathcal{A}$, $A \sqcup 1 = 1$ and $A \sqcap 0 = 0$.

Proof:

$1 = A \sqcup \sim A$ (completion law)

$= A \sqcup (\sim A \sqcap 1)$ (identity law)

$= (A \sqcup \sim A) \sqcap (A \sqcup 1)$ (distributive law)

$= 1 \sqcap (A \sqcup 1)$ (completion law)

$= (A \sqcup 1) \sqcap 1$ (commutative law)

$= A \sqcup 1$ (identity law).

By the duality principle, we also have $A \sqcap 0 = 0$■

Theorem 7-5 (Absorption laws) For $A, B \in \mathcal{A}$ (Boolean algebra), $A \sqcap (A \sqcup B) = A$ and $A \sqcup (A \sqcap B) = A$.

Proof:

$A \sqcap (A \sqcup B) = (\mathrm{A} \sqcup \mathrm{B}) \sqcap A$ (commutative law)

$= (\mathrm{A} \sqcup \mathrm{B}) \sqcap (A \sqcup 0)$ (identity law)

$= A \sqcup (B \sqcap 0)$ (distributive law in reverse)

$= A \sqcup 0$ (domination law in Theorem 7-4)

$= A$ (identity law)

By the duality principle, we also have $A \sqcup (A \sqcap B) = A$■

The next two theorems are useful in simplifying expressions. They effectively allow for the cancellation of terms on either side of an equation (similar to elementary high school algebra).

Theorem 7-6 Let $\mathcal{A}$ be a Boolean algebra and let $A, B, C \in \mathcal{A}$. If $A \sqcup C = B \sqcup C$ and $A \sqcap C = B \sqcap C$, then $A = B$.

Proof:

$A = A \sqcup (A \sqcap C)$ (absorption law)

$= A \sqcup (B \sqcap C)$ (by assumption in the theorem statement)

$= (A \sqcup B) \sqcap (A \sqcup C)$ (distributive law)

$= (A \sqcup B) \sqcap (B \sqcup C)$ (by assumption in the theorem statement)

$= (B \sqcup A) \sqcap (B \sqcup C)$ (commutative law)

$= B \sqcup (A \sqcap C)$ (distributive law in reverse)

$= B \sqcup (B \sqcap C)$ (by assumption in the theorem statement)

$= B$ (absorption law)■

Theorem 7-7 Let $\mathcal{A}$ be a Boolean algebra and let $A, B, C \in \mathcal{A}$. If $A \sqcup C = B \sqcup C$ and $A \sqcup {\sim}C = B \sqcup {\sim}C$ then $A = B$.

Proof:

$A = A \sqcup 0$ (identity law)

$= A \sqcup (C \sqcap {\sim}C)$ (completion law)

$= (A \sqcup C) \sqcap (A \sqcup {\sim}C)$ (distributive law)

$= (B \sqcup C) \sqcap (A \sqcup {\sim}C)$ (by assumption in the theorem statement)

$= (B \sqcup C) \sqcap (B \sqcup {\sim}C)$ (by assumption in the theorem statement)

$= B \sqcup (C \sqcap {\sim}C)$ (distributive law in reverse)

$= B \sqcup 0$ (completion law)

$= B$ (identity law)■

Theorem 7-8 (Associative laws) For all $A, B, C \in \mathcal{A}$ (Boolean algebra), $A \sqcup (B \sqcup C) = (A \sqcup B) \sqcup C$ and $A \sqcap (B \sqcap C) = (A \sqcap B) \sqcap C$.

Proof: The first associative law is proven below. The second associative law follows by the duality principle.

Let $X = \text{A} \sqcup (\text{B} \sqcup \text{C})$ and $Y = (\text{A} \sqcup \text{B}) \sqcup \text{C}$. If we can show that $A \sqcap X = A \sqcap Y$ and $\sim A \sqcap X = \sim A \sqcap Y$, then from the dual of Theorem 7-7, we have that $X = Y$.

We first show that $A \sqcap X = A \sqcap Y$.

$A \sqcap X = A \sqcap [A \sqcup (B \sqcup C)]$ (definition of X)

$= (A \sqcap A) \sqcup [\, A \sqcup (B \sqcup C)]$ (distributive law)

$= A \sqcup [A \sqcap (B \sqcup C)]$ (idempotent law)

$= A$ (absorption law)

On the other hand,

$A \sqcap Y = A \sqcap [(\text{A} \sqcup \text{B}) \sqcup \text{C}]$ (definition of Y)

$= [A \sqcap (A \sqcup B)] \sqcup (A \sqcap C)$ (distributive law)

$= A \sqcup (A \sqcap C)$ (absorption law on the terms in square brackets)

$= A$ (absorption law)

Thus, $A \sqcap X = A \sqcap Y$.

Next, we show that $\sim A \sqcap X = \sim A \sqcap Y$

$\sim A \sqcap X = \sim A \sqcap [A \sqcup (B \sqcup C)]$ (definition of X)

$= (\sim A \sqcap A) \sqcup [\sim A \sqcap (B \sqcup C)]$ (distributive law)

$= 0 \sqcup [\sim A \sqcap (B \sqcup C)]$ (completion law)

$= \sim A \sqcap (B \sqcup C)$ (identity)

On the other hand,

$\sim A \sqcap Y = \sim A \sqcap [(\text{A} \sqcup \text{B}) \sqcup \text{C}]$ (definition of Y)

$= [\sim A \sqcap (A \sqcup B)] \sqcup (\sim A \sqcap C)$ (distributive law)

$= [(\sim A \sqcap A) \sqcup (\sim A \sqcap B)] \sqcup \ (\sim A \sqcap C)$ (distributive law)

$= [0 \sqcup (\sim A \sqcap B)] \sqcup \ (\sim A \sqcap C)$ (completion law)

$= (\sim A \sqcap B) \sqcup (\sim A \sqcap C)$ (identity law)

$= \sim A \sqcap (B \sqcup C)$ (distributive law in reverse)

Thus, $\sim A \sqcap X = \sim A \sqcap Y$■

[Author's Remark: If you got through the previous proof and filled-in the missing details, you're really doing well. If not, don't feel bad. It is a complex proof.**]**

We conclude this sub-section by noting that De Morgan's laws also hold for Boolean algebras.

Theorem 7-9 (De Morgan's laws) For any Boolean algebra $\mathcal{A}$ and $A_1, A_2, \ldots, A_n \in \mathcal{A}$,

$$\sim(\mathrm{A}_1 \sqcup \mathrm{A}_2 \sqcup \ldots \sqcup \mathrm{A}_\mathrm{n}) = \sim\mathrm{A}_1 \sqcap \sim\mathrm{A}_2 \sqcap \ldots \sqcap \sim\mathrm{A}_\mathrm{n}$$
$$\sim(\mathrm{A}_1 \sqcap \mathrm{A}_2 \sqcap \ldots \sqcap \mathrm{A}_\mathrm{n}) = \sim\mathrm{A}_1 \sqcup \sim\mathrm{A}_2 \sqcup \ldots \sqcup \sim\mathrm{A}_\mathrm{n}.$$

7.3 Examples

As noted at the beginning of this section, propositional logic and set operations are examples of Boolean algebras. The mappings between Boolean algebra operations and special elements to propositional logic and set operations are shown in Table 19.

Table 19. Comparison of Terms from Boolean Algebra, Logic and Set Theory

Boolean Algebra	Propositional Logic	Set Theory
$A \sqcap B$ (meet)	$A \wedge B$ (conjunction)	$A \cap B$ (intersection)
$A \sqcup B$ (join)	$A \vee B$ (disjunction)	$A \cup B$ (union)
$\sim A$ (completion)	$\neg A$ (negation)	$\Omega - A$ or just $-A$ (complement)
0, 1 (identity elements)	F (false), T(true)	ϕ (empty set), Ω (universal set)

In terms of sets, there are actually many Boolean algebras. To form a Boolean algebra, select a universal set (domain of interest), e.g., the set of natural numbers, and then take the set of all subsets (the power set). This approach also allows for finite Boolean algebras. For example, let $\Omega = \{a, b, c\}$ then we can form a Boolean algebra under set intersection and union where the elements are from the power set of Ω, i.e., $\phi, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}$.

As another example, consider all the possible statements that are equivalent to a given statement A. In notation, this is represented as $\|A\|$. For example, $1 = A \cup \sim A = A \cup (\sim A \sqcap 1) = (A \sqcup \sim A) \sqcap (A \sqcup 1) = 1 \sqcap (A \sqcup 1)$ and so, $A \cup \sim A, A \cup (\sim A \sqcap 1), (A \sqcup \sim A) \sqcap (A \sqcup 1)$ and $1 \sqcap (A \sqcup 1)$ are all represented by $\|1\|$. Of course, $\|A\|$ represents an infinite number of statements. Now, let $\mathcal{B}$ be the collection of all elements of the form $\|A\|$ for some logic statement A. Meet and join for $\mathcal{B}$ are defined using conjunction and disjunction from propositional logic:

$$\|A\| \sqcup \|B\| = \|A \vee B\|$$
$$\|A\| \sqcap \|B\| = \|A \wedge B\|$$

Further, let $0 = \|F\|$ and $1 = \|T\|$, and let $\sim\|A\| = \|\neg A\|$.

It is left as an exercise for the reader to show that $\mathcal{B}$ is a Boolean algebra. ***Hint:*** *Show that the axioms for a Boolean algebra hold true. This task reduces to making use of the existing properties of propositional logic.*

7.4 Switching Circuits

Switching circuits can be modeled using Boolean algebra where meet ($\sqcap$) represents switches in series and join ($\sqcup$) represents parallel switching elements. The state of a switch (as represented by an element in a Boolean algebra) is true (1) if the switch is closed (i.e., current can flow) and false (0) if it is open (i.e., current cannot flow).

For example, consider the representation of the series and parallel switching circuits in Figure 5. A and B represent switches (shown as open in the figure). The diamond shaped points are inputs and outputs for electrical current. In the series arrangement, current flows only when both A and B are closed. In the parallel arrangement, current flows if either A or B is closed.

In the series arrangement in Figure 5, if A and B have the value 1, then $A \sqcap B = 1$ and current can flow. In the parallel arrangement, if either A or B has the value 1, then $A \sqcup B = 1$ and current can flow.

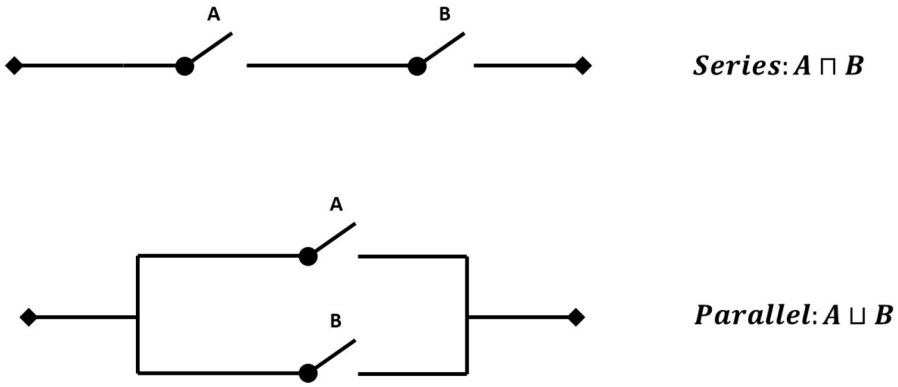


Figure 5. Series and Parallel Circuits

Figure 6 illustrates two additional points:

- If several switches are labelled with the same variable, they are all open or all closed (see the switches labelled as B in the figure).
- As shown in the figure, a switch labelled as A and its completion $\sim A$ are in opposite states (one closed and the other open).

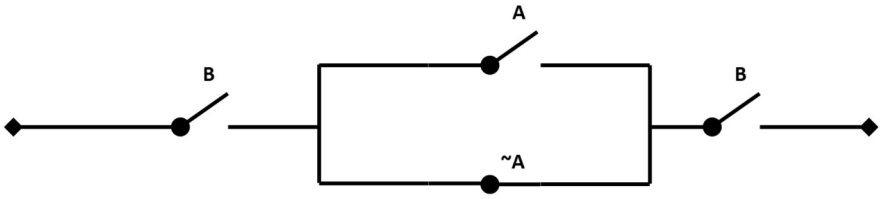


Figure 6. Synchronized Switches

The circuit in Figure 6 can be represented as $B \sqcap (A \sqcup \sim A) \sqcap B = B \sqcap 1 \sqcap B = B$, which illustrates a key point, i.e., it is possible to simplify a switching circuit by representing the circuit in Boolean algebra notation and then simplifying the expression to arrive at a simpler circuit. The circuit in Figure 6 can be simplified to the circuit consisting of only one instance of switch B.

As another example of circuit simplification using Boolean algebra, consider the circuit in Figure 7.

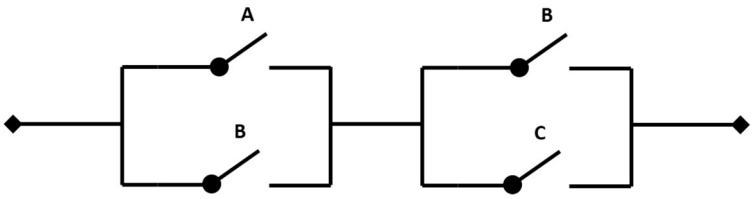


Figure 7. Circuit Simplification Example

The circuit can be represented as $(A \sqcup B) \sqcap (B \sqcup C)$ which can be simplified as follows:

$(A \sqcup B) \sqcap (B \sqcup C)$

$= (B \sqcup A) \sqcap (B \sqcup C)$ (commutative law)

$= [B \sqcap (B \sqcup C)] \sqcup [A \sqcap (B \sqcup C)]$ (distributive law)

$= B \sqcup [(A \sqcap B) \sqcup (A \sqcap C)]$ (absorption law on the left and distribution law on the right-side of the previous statement)

$= [(B \sqcup (A \sqcap B)] \sqcup (A \sqcap C)$ (associative law)

$= B \sqcup (A \sqcap C)$ (absorption law on the left-side of the above).

The simplified circuit is shown in Figure 8.

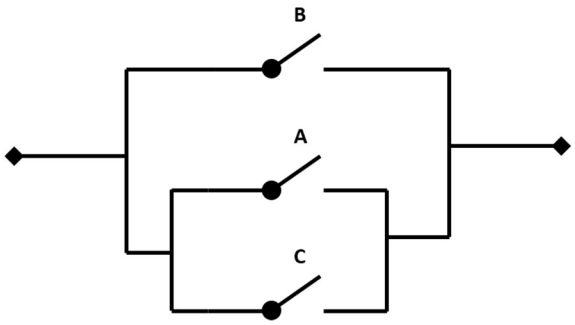


Figure 8. Simplified Circuit

The simplification is not unique. For example, $(A \sqcup B) \sqcap (B \sqcup C)$ can be written in several reduced forms:

- using NANDs: $(A \uparrow C) \uparrow \sim B$
- using joins and completions: $\sim(\sim A \sqcup \sim C) \sqcup B$
- using implies and completions: $(A \Rightarrow \sim C) \Rightarrow B$.

The previous examples can all be classified as series-parallel circuits.

It is possible to apply Boolean algebra to more complex circuits such as the bridge circuit shown in Figure 9.

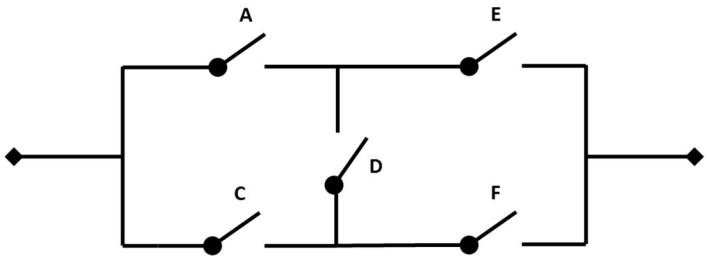


Figure 9. Bridge Circuit

There are several approaches for describing the bridge circuit in Figure 9 as well as other more complex circuits.

Path Approach: In this approach, one determines all combinations of closed switches that allow current to flow from one terminal to another. For the bridge circuit in Figure 9, the possible paths are AE, CF, ADF and CDE. In terms of a Boolean algebra statement, this can be expressed as $(A \sqcap E) \sqcup (C \sqcap F) \sqcup (A \sqcap D \sqcap F) \sqcup (C \sqcap D \sqcap E)$.

Breaks Approach: In this approach, one determines all possible combinations of switch openings that prevent current flow between the terminals. If A and C are open, the circuit is broken; or if E and F are open, the circuit is broken; or if A, D and F are open, the circuit is broken; or if C, D and E are open, the circuit is broken. See the illustration in Figure 10.

So, the condition for the circuit being open is

$$(\sim A \sqcap \sim C) \sqcup (\sim E \sqcap \sim F) \sqcup (\sim A \sqcap \sim D \sqcap \sim F) \sqcup (\sim C \sqcap \sim D \sqcap \sim E)$$

The condition for the circuit being closed is the negation of the above statement which reduces to the following by application of De Morgan's law

$$(A \sqcup C) \sqcap (E \sqcup F) \sqcap (A \sqcup D \sqcup F) \sqcap (C \sqcup D \sqcup E)$$

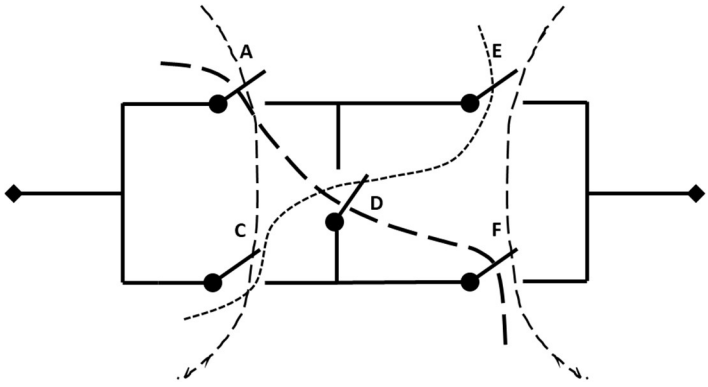


Figure 10. Breaks Approach

For a more complete discussion of the many facets of Boolean algebra as applied to circuit design, see the books by Solomon [30] and Whitesitt [31].

7.5 Exercises

1. Show that the completion of an element in a Boolean algebra $\mathcal{A}$ is unique. ***Hint**: Assume that X and Y are completions of $A \in \mathcal{A}$. Then by the definition of completion, $A \sqcup X = 1, A \sqcap X = 0$ and $A \sqcup Y = 1, A \sqcap Y = 0$. Thus, $A \sqcup X = A \sqcup Y$ and $A \sqcap X = A \sqcap Y$. From Theorem 7-6, the result follows.*
2. For a Boolean algebra, show that it is possible for $A \sqcap B = A \sqcap C$ and yet $B \neq C$. ***Hint**: Consider a Boolean algebra over the set of natural numbers.*
3. Prove De Morgan's laws for two elements of a Boolean algebra, i.e., prove $\sim(A \sqcup B) = \sim A \sqcap \sim B$ and $\sim(A \sqcap B) = \sim A \sqcup \sim B$. ***Hint**: We just need to prove the first identity and the second follows from the duality principle. For the first identity, we need to show that $\sim A \sqcap \sim B$ fulfills the definition of completion for $A \sqcup B$, i.e., show that $(A \sqcup B) \sqcup (\sim A \sqcap \sim B) = 1$ and $(A \sqcup B) \sqcap (\sim A \sqcap \sim B) = 0$. To prove the former identity in the previous sentence, use the distributive law, and then the associative law along with the completion and domination laws to get $(A \sqcup B) \sqcup (\sim A \sqcap \sim B) = [(A \sqcup B) \sqcup \sim A] \sqcap [(A \sqcup B) \sqcup \sim B] = 1 \sqcup 1 = 1$. The proof that $(A \sqcup B) \sqcap (\sim A \sqcap \sim B) = 0$ is similar.*
4. Write down a Boolean statement corresponding to the following circuit:

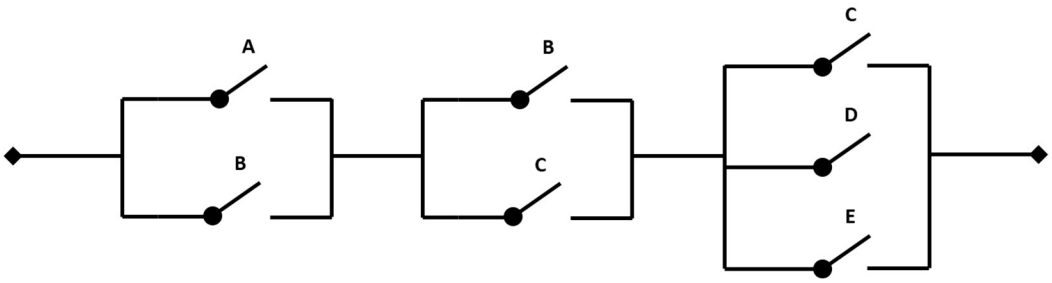


Figure 11. Series-Parallel Circuit

5. Draw a circuit corresponding to the Boolean statement $A \sqcap (B \sqcup C \sqcup D) \sqcap (E \sqcup F)$.

8 Functions

8.1 Terminology and Examples

A **function** is a relation between sets A and B that associates to every element of A exactly one element in set B. The set A (mapping from) is called the **domain** of the function and the set B (mapping to) is called the **codomain** of the function.

In terms of notation, a function from set A to set B is represented as $f: A \to B$.

Figure 12 shows several example functions and several non-functions. The more general term "**mapping**" is used to describe both functions and non-functions.

- The mapping f is a function since it maps each element of A to exactly one element in B. Nothing is mapped to 3 but that doesn't violate the definition of a function.
- The mapping g is not a function since c is mapped to two elements in B.
- The mapping h is not a function since b is not mapped to anything in B.
- The mapping j is a function since each element of A is mapped to exactly one element of B. It is allowed for several elements of the domain (b and c in this case) to be mapped to one element in the codomain (2 in this case).

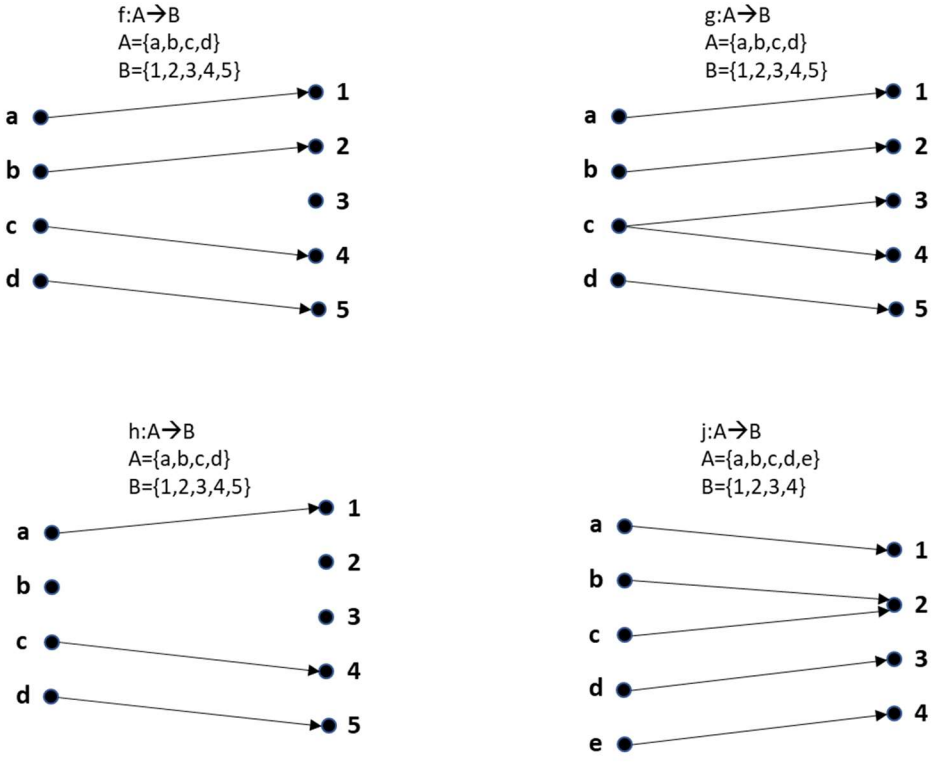


Figure 12. Example Functions and Non-functions

The above examples involve finite domains and codomains but it is possible to have functions between infinite sets. For example, the mapping of each real number x to x^2 is a function. This is

written as $f(x) = x^2$, where the domain is the set of all real numbers (represented by the symbol $\mathbb{R}$) and the codomain is the set of all positive real numbers (noting that the square of a negative number is positive). For example, and to illustrate the above notation for f, $f(3) = 3^2 = 9$ simply means that given the element 3 from the domain, square the number to get the mapping to the codomain (in this case, 9).

As another example, recall mapping from $\mathbb{Z}\ to\ \mathbb{N}$ that we presented in Table 14 of Section 6.2. The domain is the set of all integers (represented as $\mathbb{Z}$) and the codomain is the set of natural numbers (represented as $\mathbb{N}$). The function can be written more formally as follows:

$$f(x) = \begin{cases} 2x + 2, if\ x \geq 0 \\ -(2x + 1), if\ x < 0 \end{cases}$$

If one reverses the direction of a mapping for a function, then the inverse is obtained. More formally, the **inverse of a function** $f: A \rightarrow B$ (denoted as $f^{-1}: B \rightarrow A$) maps $y \in B$ to $x \in A$ such that $y = f(x)$. The inverse may or may not be a function. For example, the inverse of the function j in Figure 12 is not a function since j^{-1} maps 2 to b and c.

Another way to view a function is as a set of pairs $\big(x, f(x)\big)$ for every x in the domain of f. Using this view, the inverse of f is just the set of transposed pairs, i.e., $(f(x), x)$.

A function $f: A -> B$ is said to be **injective** (also known as "one-to-one") if $f(x) \neq f(y)$ for any two different elements x and y of A. This is an additional requirement beyond just being a function. Equivalently, f is injective if $\forall y \in B, f^{-1}(y)$ maps to exactly one element in A.

- For example, the function j in Figure 12 is not injective since $j^{-1}(2)$ is mapped to b and c.
- The function $p(x) = x^2$ is not injective since $p^{-1}(y)$ is mapped to $-\sqrt{y}$ and $\sqrt{y}$, e.g., $p^{-1}(4)$ is mapped to -2 and 2.
- However, $q(x) = x^3$, where $x \in \mathbb{R}$, is injective. The reason is that $q^{-1}(y) = \sqrt[3]{y}$ (cubed root of y) has only one solution.

A function $f: A -> B$ is said to be **surjective** (also known as "onto") if $f^{-1}(y)$ is defined for every $y \in B$. For example, the function f in Figure 12 is not surjective since $f^{-1}(3)$ is undefined.

A function that is both injective and surjective is said to be **bijective**.

Figure 13 may help the reader visualize the difference between surjective and injective. On the left of the figure, different elements in A are mapped to different elements of B, but there is no mapping from A to some of the elements in B. So, function f is injective but not surjective. On the right, function g is not injective since it maps different elements of A to the same element of B, but g is surjective since for every element of B then is an element of A that is mapped to it.

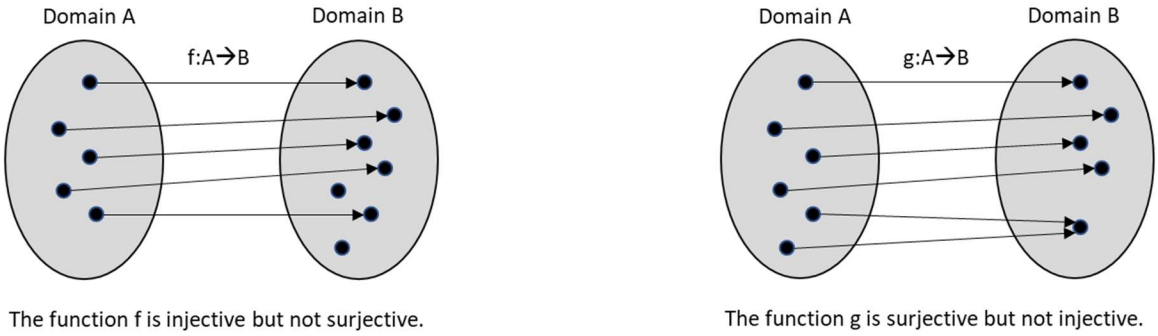


Figure 13. Injective and Surjective Functions

Function h in Figure 14 is bijective, since h maps every element in A to a unique element in B and h^{-1} maps every element of B to a unique element in A.

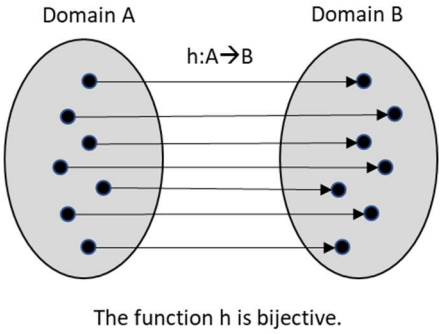


Figure 14. Bijective Function

The subset of the codomain to which the domain of a function $f: A \to B$ is mapped is called **image** or **range** of f and is represented as $f(A)$. Note that $f(A) \subset B$. If $f(A) = B$, the function f is surjective.

8.2 Composition of Functions

A function can take another function as its argument. For example, take $f(x) = x^2$ and $g(x) = x - 2$. The composition of f with g (written as $f \circ g$) and is defined as $f\big(g(x)\big) = [g(x)]^2 = (x-2)^2$. For this example, $f: \mathbb{R} \to \mathbb{R}^+$ where $\mathbb{R}^+$ stands for the non-negative real-numbers, and $g: \mathbb{R} \to \mathbb{R}$ which implies that $f \circ g: \mathbb{R} \to \mathbb{R}^+$.

In general, if one is to take the composition of two functions, the codomain of one function needs to match the domain of the other. For example, if $f: B \to C$ and $g: A \to B$, then $f \circ g: A \to C$. Figure 15 depicts the composition of two functions with finite domains and codomains. It is necessary condition for composition that the codomain of g at least be a subset of the domain of f.

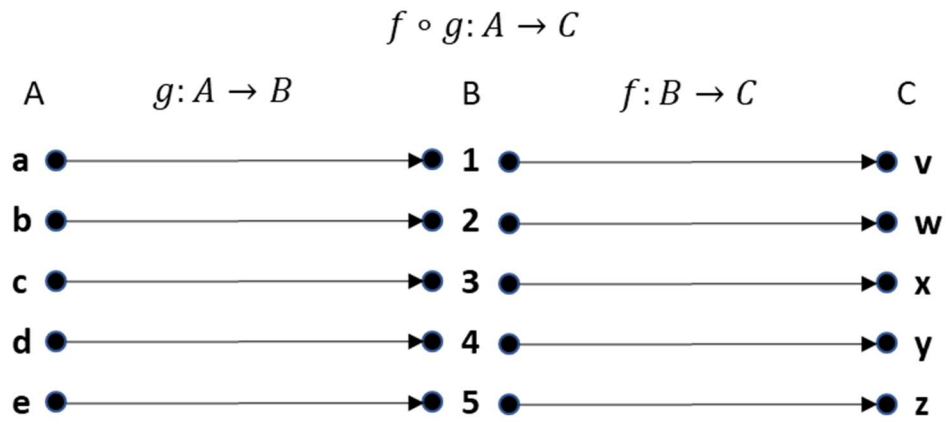


Figure 15. Example of Composition of Functions

8.3 Visual Representations

There are various methods for visually representing a function. We have already seen the use of a table to represent a function that maps the integers to the natural numbers, see Table 14 of Section 6.2. Another example is shown in Table 20 which represents the function $f(n) = 3n + 1$ where $n \in \mathbb{N}$.

Table 20. Mapping of $\mathbb{N}$ to integers of the form $3n + 1$

n	0	1	2	3	4	5	6	…
$f(n)$	1	4	7	10	13	16	19	…

A table works well for functions that map between countable (finite or infinite) sets. A bar graph can also be used for mappings between countable (finite or infinite) sets. In Figure 16, a bar chart is used to represent $f(n) = 3n + 1$ where $n \in \mathbb{N}$. The chart was generated using Microsoft Excel.

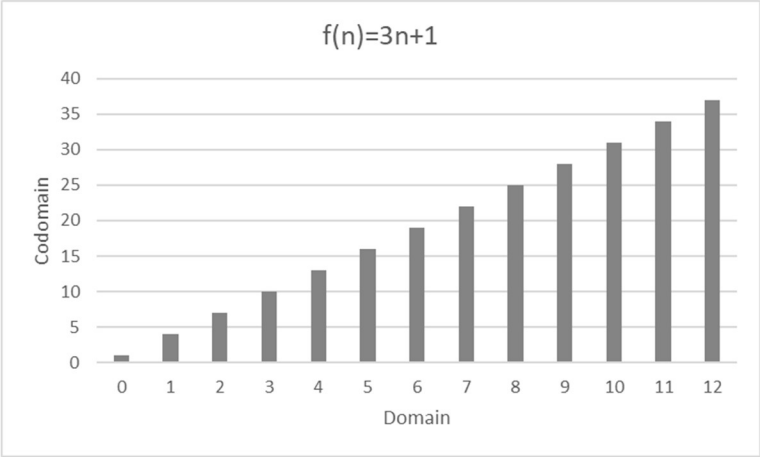


Figure 16. Bar Chart for f(n)=3n+1

For functions that map between uncountable sets, a continuous graph is typically used. Figure 17 show the graph of $f(x) = x^2 - 2$ where $x \in \mathbb{R}$. The horizontal axis (usually called the x-axis) represents the values in the domain of f, and the vertical axis (usually called the y-axis) represents the mapping by f. For example, $f(3) = 3^2 - 2 = 7$ is represented by the point $(3,7)$ on the graph.

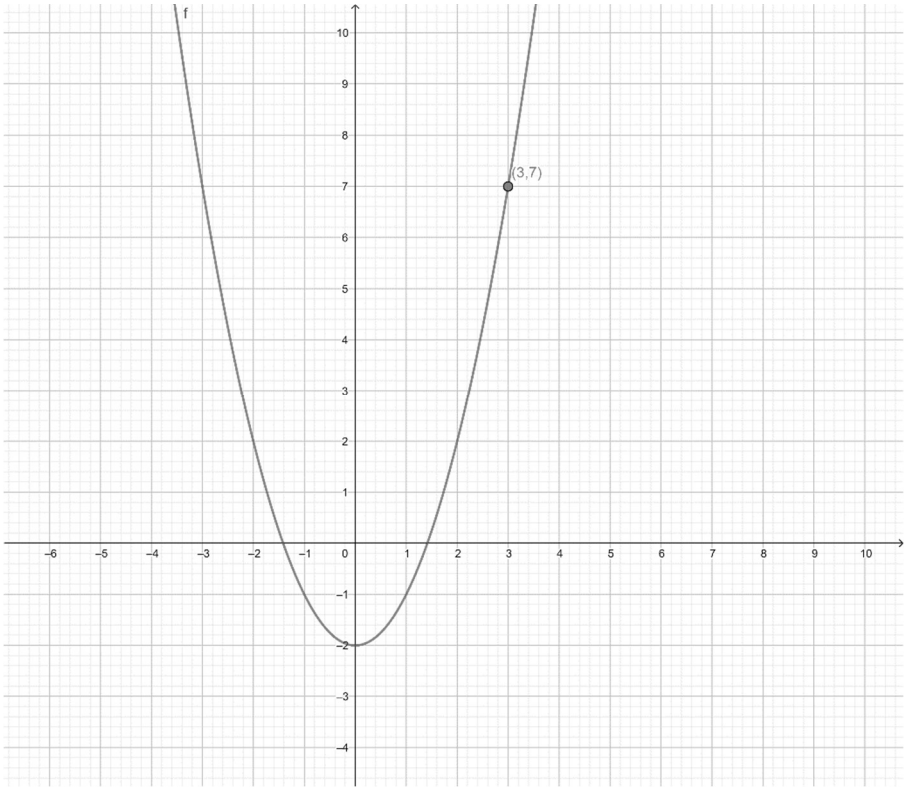


Figure 17. Graph of f(x)

8.4 Absolute Value

The absolute value of a variable x is defined as follows:

$$f(x) = |x| = \begin{cases} x, & if\ x \geq 0 \\ -x, & if\ x < 0 \end{cases}$$

Figure 18 shows the graph of $f(x) = |x|$.

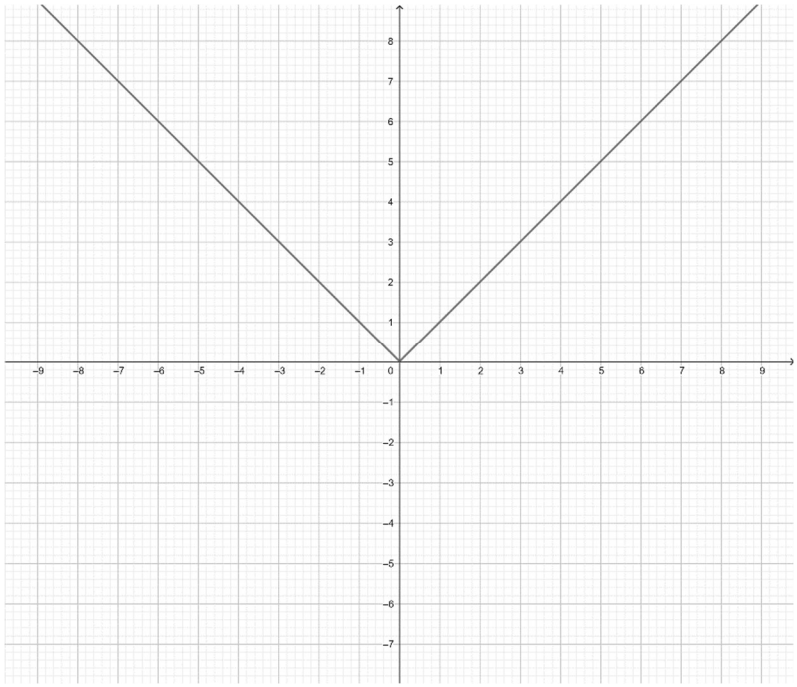


Figure 18. Graph of the Absolute Value of x

8.5 Polynomials

A **polynomial** of one variable is an expression involving the sum of powers of a variable where each term is multiplied by a constant. The general form of a polynomial is

$$f(x) = a_r x^r + a_{r-1} x^{r-1} + \cdots + a_2 x^2 + a_1 x^1 + a_0.$$

The a_i terms are called **coefficients**. The highest power of the variable in a polynomial is called the **degree** of the polynomial. For example, the function $g(x) = x^3 + 3x^2 - x - 2$ is of degree 3 and the coefficients are 1, 3, -1 and -2 (going from the highest power of x to the lowest). For $x \in \mathbb{R}$, the graph of $g(x)$ is shown in Figure 19.

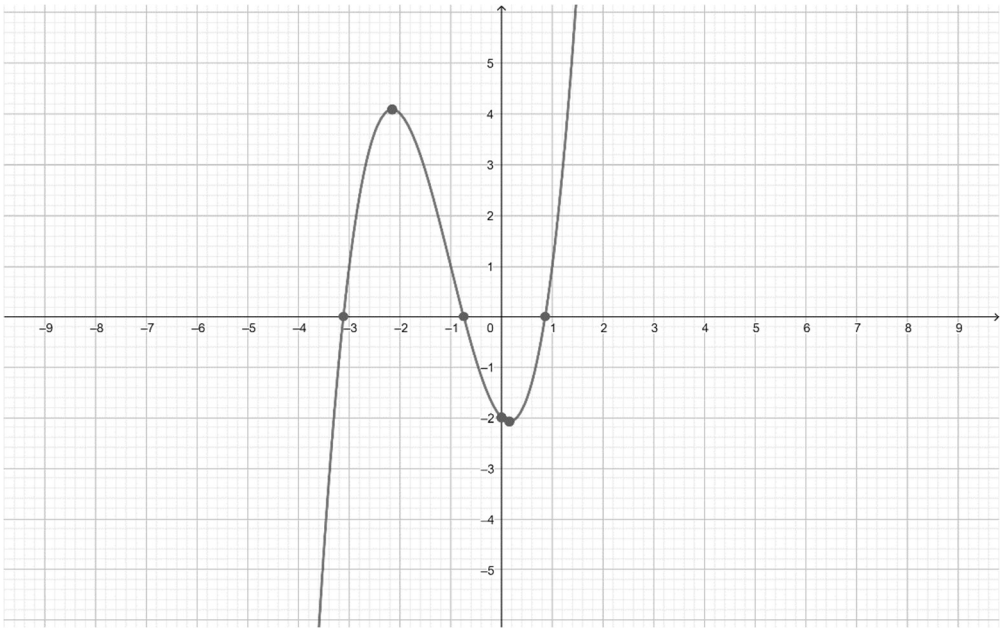


Figure 19. Graph of a Cubic Polynomial

It is also possible to apply the absolute value function to another function. Figure 20 shows the graph of $h(x) = |g(x)| = |x^3 + 3x^2 - x - 2\ |$. As can be seen, this is almost the same as g(x), except the part of the graph below the x-axis is reflected to the positive-side of the x-axis.

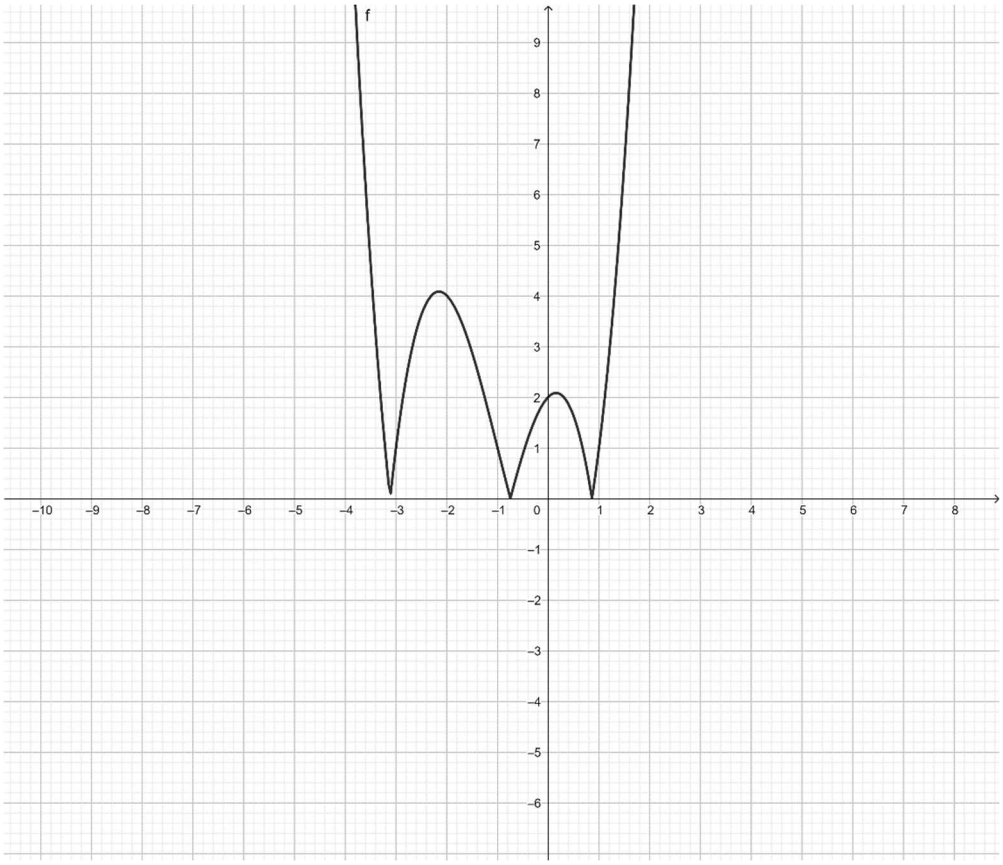


Figure 20. Graph of the Absolute Value of a Cubic Polynomial

Polynomials have been studied extensively in Mathematics. For a summary of the many properties of polynomials see the Wikipedia articles on the topic [32] [33].

8.6 Exponential and Logarithmic Functions

An exponential function is simply a real number raised to the power of a variable, e.g., $f(x) = 3^x$. The general form is $f(x) = a^x$, where a is a constant. Table 21 shows some sample values of 3^x. As x gets larger, 3^x becomes large very quickly – in fact, faster than any polynomial. As x gets smaller, 3^x approaches zero.

Table 21. Some values of the exponential function with $a = 3$

x	$-\infty$	…	-4		-3	-2	-1	0	1	2	3	4	…	∞
3^x	0	…	$3^{-4} = \frac{1}{81}$		$3^{-3} = \frac{1}{27}$	$3^{-2} = \frac{1}{9}$	$3^{-1} = \frac{1}{3}$	1	3	9	27	81	…	∞

Figure 21 depicts a graph of the exponential function $f(x) = 3^x$ (higher up and to the left) and its inverse (lower and to the right). For negative values of x it looks like 3^x equals zero after about $x = -4$ but in fact, it is not zero (just very small and beyond the granularity of the graph).

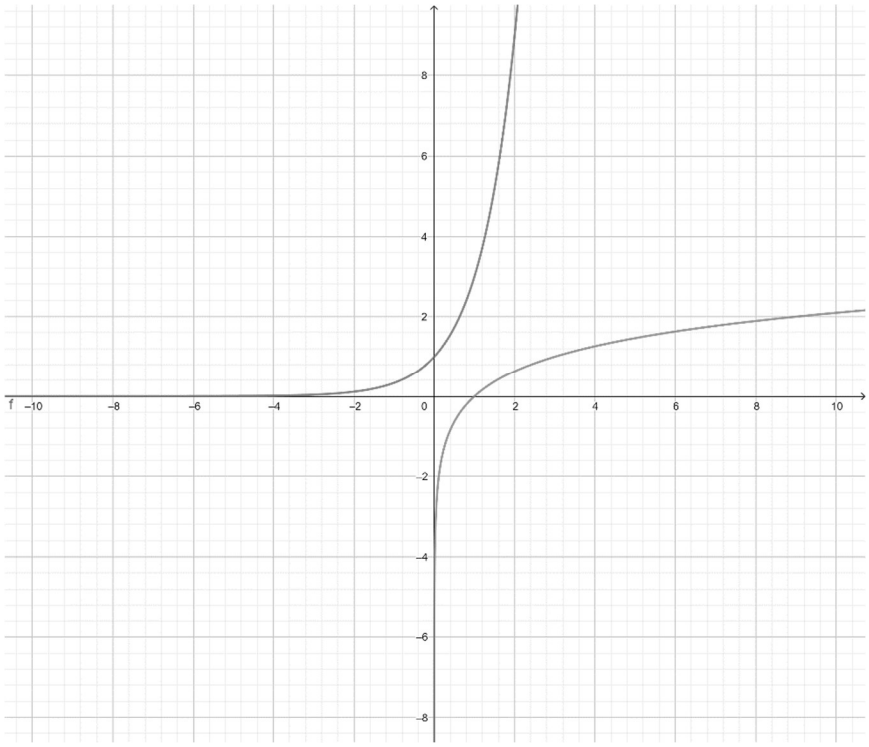


Figure 21. Graph of Exponential Function with $a = 3$ and its Inverse

The inverse function for a given exponential function a^x is called the logarithm base a which is written as $\log_a x$. The logarithm of a given number x is the exponent to which another fixed number, the base a, must be raised, to produce that number x. Saying the same thing in more formal notation

$$\log_a x = y \Leftrightarrow a^y = x$$

For example, $\log_3 81 = 4$ since 3 needs to be raised to the 4 power to equal 81. In Figure 21, the function $\log_3 x$ is shown in red.

Recall from high school algebra that $a^x a^y = a^{x+y}$, $\frac{a^x}{a^y} = a^{x-y}$ and $(a^x)^y = a^{xy}$. Product, division and power rules can also be shown for logarithms.

Theorem 8-1 The following rules hold true for logarithms:

- $\log_a(x\,y) = \log_a x + \log_a y$ (Product)
- $\log_a\left(\frac{x}{y}\right) = \log_a x - \log_a y$ (Division)
- $\log_a x^r = r \log_a x$ (Power)
- $\log_a \sqrt[r]{x} = \frac{1}{r}\log_a x$ (Root)

Proof: (Product) Let $w = \log_a(x)$ and $z = \log_a(y)$. Then by definition of a logarithm, $a^w = x$ and $a^z = y$. Multiplying the two equations gives $a^w a^z = a^{w+z} = xy$ which implies (by the definition of logarithm) that $w + z = \log_a(x\,y)$ but $w + z = \log_a x + \log_a y$.

(Power) Let $z = \log_a(x^r)$ which implies, by the definition of a logarithm, that $a^z = x^r$. Next, take the r^{th} root of each side of the equation to get $a^{\frac{z}{r}} = x$. Using the definition of logarithm, we get $\frac{z}{r} = \log_a x$ which implies the desired result, i.e., $z = r \log_a x$.

(Division) This follows from the product and power properties:

$$\log_a\left(\frac{x}{y}\right) = \log_a(xy^{-1}) = \log_a x + \log_a y^{-1} = \log_a x - \log_a y$$

(Root) Since the power property holds for any $r \in \mathbb{R}$ and $\sqrt[r]{x} = x^{\frac{1}{r}}$, the root property follows from the power property■

8.7 Transforming the Graph of a Function

Once one knows the graph of a particular function, it is possible to determine several related graphs via transformation of the original graph.

8.7.1 Up and Down

By adding or subtracting a constant from a function, the associated graph is moved up or down by the value of the constant.

For example, consider $f(x) = x^3 - 2x$ (the middle graph in Figure 22).

- If 3 is added to $f(x)$, we get the top graph $g(x) = x^3 - 2x + 3$. For a given value of x, $g(x)$ is 3 units above $f(x)$.
- Similarly, 3 can be subtracted from $f(x)$ to get $h(x) = x^3 - 2x - 3$ (the bottom graph in the figure).

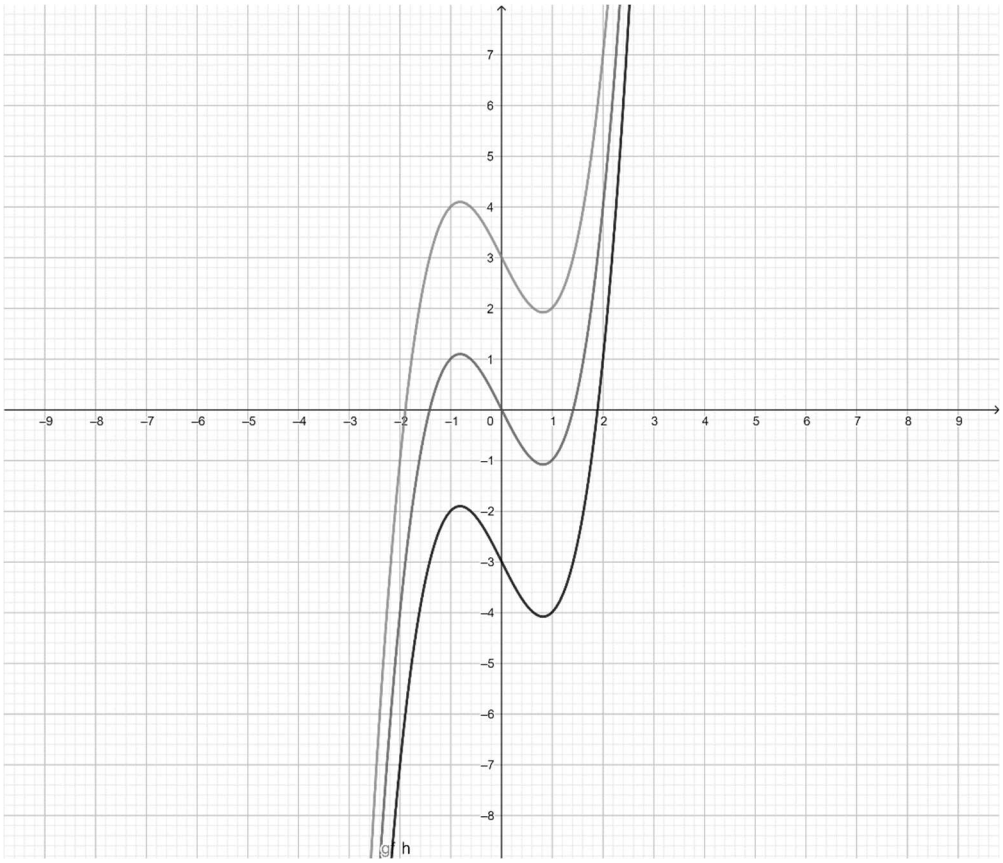


Figure 22. Vertical Transformation

8.7.2 Left and Right

It is also possible to transform a graph by moving it to the left or right. This is done by replacing x with $x + a$ where a is a constant.

Again, consider $f(x) = x^3 - 2x$ (the middle graph in Figure 23). Replace x with $x - 2$ to get the right graph $g(x) = f(x - 2) = (x - 2)^3 - 2(x - 2)$. (This is the composition of $f(x)$ with the function $p(x) = x - 2$, i.e., $g = f \circ p$.) The graph of $g(x)$ is same as $f(x)$ except moved two units to the right.

Similarly, if we replace x by $x + 2$ in $f(x)$, we get $h(x) = f(x + 2) = (x + 2)^3 - 2(x + 2)$ which is the left graph in the figure. This transformation moves $f(x)$ to units to the left.

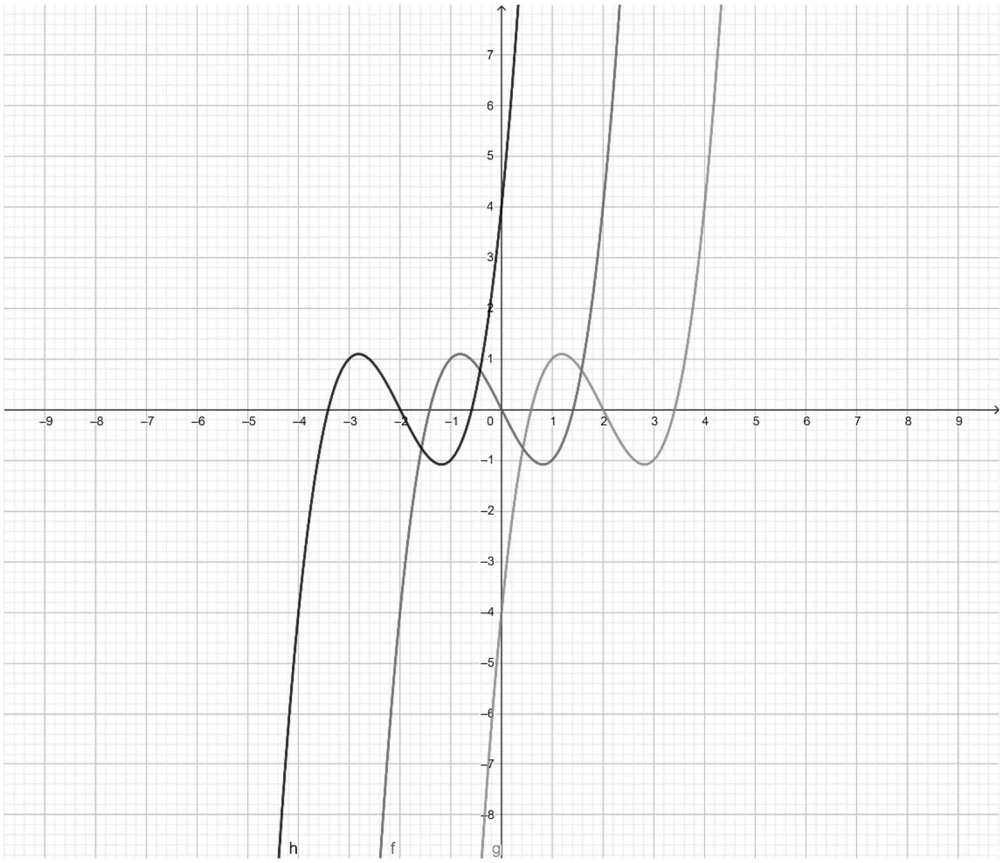


Figure 23. Horizontal Transformation

It is possible to do both a vertical and horizontal transformation. For example, the function $k(x) = f(x - a) + b$ moves the function $f(x)$ up b units and to the right by a units, assuming $a > 0$ and $b > 0$.

8.7.3 Reflections

If one multiplies a function by -1, the graph is reflected about the horizontal axis. Figure 24 shows the graph $f(x) = 2^x$ and the reflection of $f(x)$ about the horizontal axis, i.e., $h(x) = -f(x) = -2^x$.

If one replaces x by $-x$, the graph of a function is reflected about the vertical axis. For example, $g(x) = f(-x) = 2^{-x}$ is the reflection of $f(x)$ about the vertical.

If one replaces x by $-x$ and multiplies a function by -1, the function is reflected about the horizontal axis and then the vertical axis. The two reflections have the same result as a 180-degree rotation. In Figure 24, $p(x) = -f(-x) = -2^{-x}$ is a 180-degree rotation of $f(x)$.

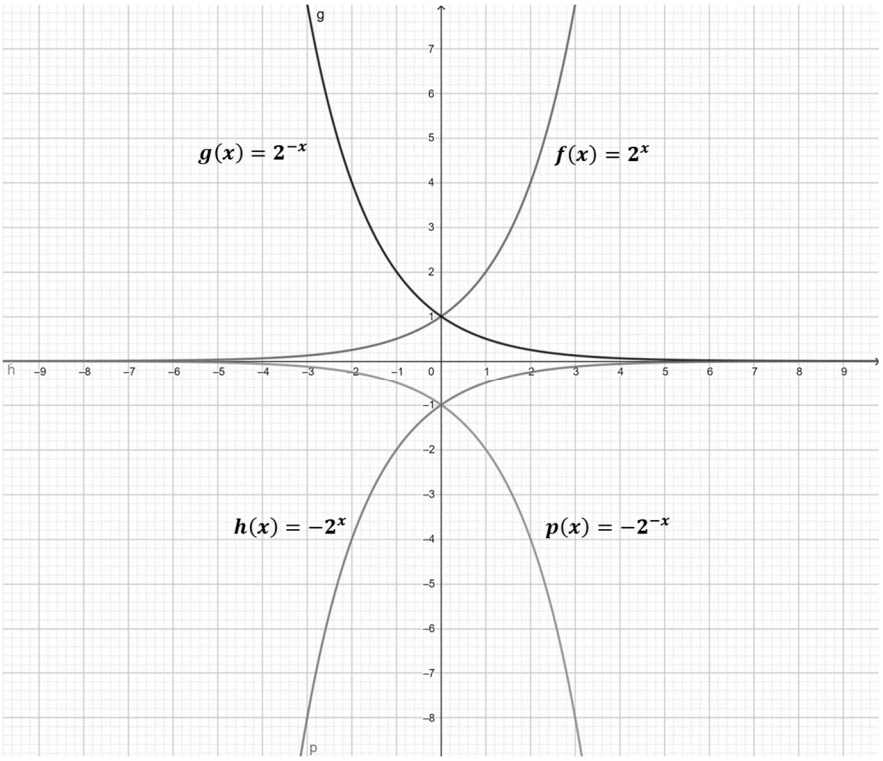


Figure 24. Reflections of an Exponential Function

8.7.4 Stretching a Graph

It is possible to stretch (expand) a graph in either the vertical or horizontal direction.

- To stretch a graph in the vertical direction, multiple by a positive constant a where $a > 1$. In Figure 25, the graph of $f(x) = x^3 - 2x$ is stretched in the vertical direction by multiplying $f(x)$ by 3 to get $g(x) = 3f(x) = 3x^3 - 6x$.
- To stretch a graph in the horizontal direction, replace x by ax where $0 < a < 1$. In Figure 25, the graph of $f(x)$ is stretched in the horizontal direction by replacing x with $\frac{x}{3}$ to get $h(x) = f\left(\frac{x}{3}\right) = \frac{1}{27}x^3 - \frac{2}{3}x$.

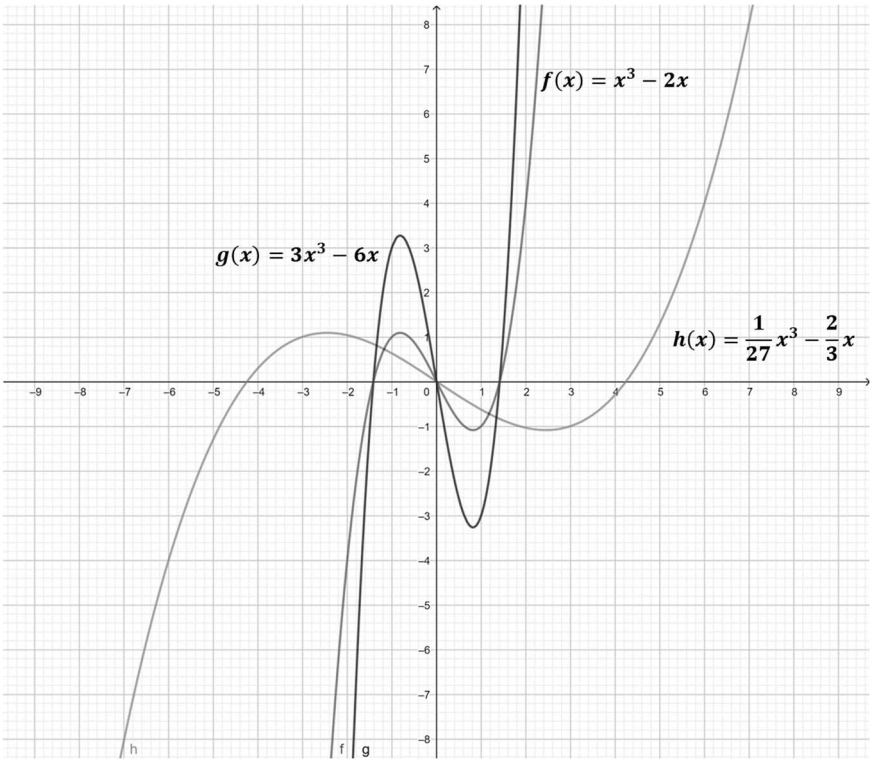


Figure 25. Stretching a Graph

8.7.5 Compressing a Graph

It is also possible to do the opposite of stretching a graph, i.e., compressing a graph.

- To compress a graph in the vertical direction, multiple by a positive constant a where $0 < a < 1$. In Figure 26, the graph of $f(x) = x^3 - 2x$ is compressed in the vertical direction by multiplying $f(x)$ by $\frac{1}{3}$ to get $g(x) = \frac{1}{3}f(x) = \frac{1}{3}x^3 - \frac{2}{3}x$.
- To compress a graph in the horizontal direction, replace x by ax where $a > 1$. In Figure 26, the graph of $f(x)$ is compressed in the horizontal direction by replacing x with $3x$ to get $h(x) = f(3x) = 27x^3 - 6x$.

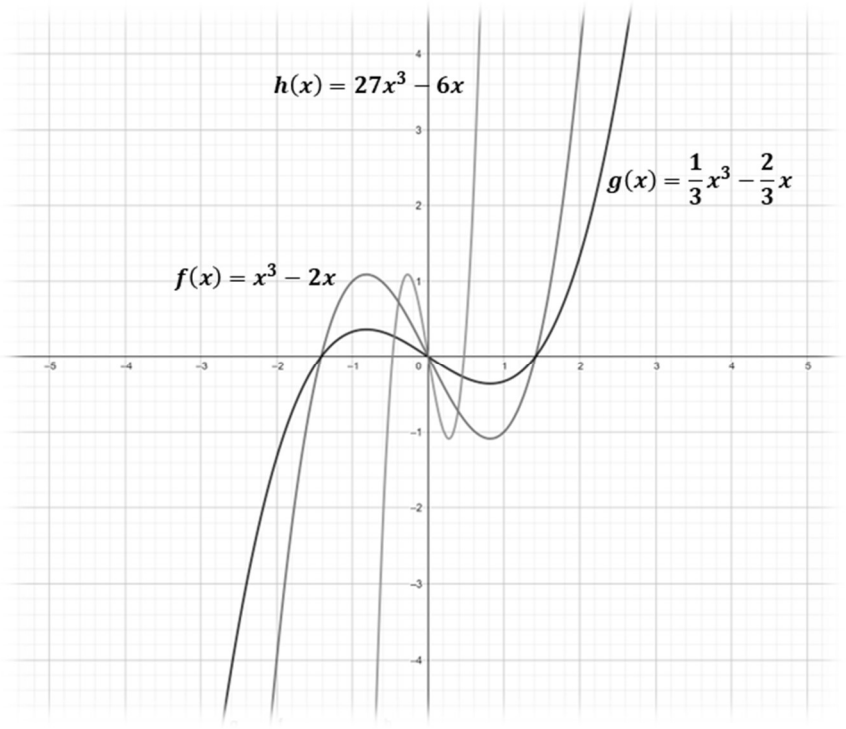


Figure 26. Compressing a Graph

8.7.6 Summary of Transformations

Table 22 provides a summary of the various function transformation discussed in the previous subsections.

Table 22. Summary of Function Transformations

Transformation	Effect
$\boldsymbol{f(x) + c}$	Moves $f(x)$ up or down c units depending on whether c is positive or negative
$\boldsymbol{f(x + c)}$	Moves $f(x)$ to the right c units if $c < 0$ Moves $f(x)$ to the left c units if $c > 0$
$\boldsymbol{f(-x)}$	Reflects $f(x)$ about the vertical axis
$\boldsymbol{-f(x)}$	Reflects $f(x)$ about the horizontal axis
$\boldsymbol{cf(x)}$	Stretches $f(x)$ in the vertical direction by a multiple of c when $c > 1$ Compresses $f(x)$ in the vertical direction by a multiple of c when $0 < c < 1$
$\boldsymbol{f(cx)}$	Stretches $f(x)$ in the horizontal direction by a factor of c when $0 < c < 1$ Compresses $f(x)$ in the horizontal direction by a factor of c when $c > 1$

8.8 Exercises

1. For the function $f(x) = x^2 - 4$ with $x \in \mathbb{R}$, what is the codomain of $f(x)$? Why is $f(x)$ not bijective?
2. What is the inverse of $f(x) = 7x + 4$? ***Hint**: Modify the equation $y = 7x + 4$ to express x in terms of y.*
3. What is the inverse of $f(x) = x^5$ given that $x \in \mathbb{R}$?
4. **Try** to find the inverse of $f(x) = 2x^4 - 5x^2 + 2x$? ***Hint**: The point of the exercise is to illustrate that in general, it is hard to find a closed expression for the inverse of a polynomial. Wolfram Alpha (https://www.wolframalpha.com) is able to find the inverse (the answer would be very hard to derive manually).*
5. As was noted previously, one way to view the function $f(x)$ is as the collection of pairs $\big(x, f(x)\big)$. $f^{-1}(x)$ can be viewed as the collection of transposed pairs, i.e., $\big(f(x), x\big)$. With this view in mind, what can be said about the relationship between $f(x)$ and $f^{-1}(x)$ in terms of a reflection? ***Hint**: Graphically $f^{-1}(x)$ is the reflection of $f(x)$ about a straight line.*
6. If $f(x) = 3^x$ and $g(x) = x^2$ what is $f \circ g(x)$? What is $f \circ g(2)$?
7. Draw the graph of $f(x) = |x^2 - 3|$?
8. Simplify the expression $\log_2[(x^3y^{-5})/z^7]$.
9. Do the following transformations for $f(x) = 2x^4 - 5x^2 + 2x$: compress in the vertical direction by a factor of $\frac{1}{2}$, compress in the horizontal direction by a factor of 3 and then reflect about the vertical axis. ***Answer**: $p(x) = 81x^4 - \frac{45}{2}x^2 - 3x$*

9 Number Theory

9.1 Background

Number theory is a branch of mathematics devoted primarily to the study of the properties of integers (and in particular, the natural numbers). Number theory is included here for two main reasons, i.e., (1) some of the basic results are needed elsewhere in this book and (2) to illustrate some techniques of proof (mathematical induction in particular).

9.1.1 Well-ordering Principle

The **well-ordering principle** states that every non-empty set S of positive integers contains a least element. In other words, there exist $x \in S$ such that $x \leq y$ for every $y \in S$. This principle, while appearing obvious, is critical in the proof of many important theorems.

[Note: The well-ordering principle does apply to sets other than the positive integers. More generally, a set X is well-ordered by a strict total order if every non-empty subset of X has a least element under the ordering. Where "strict total order" entails a binary operation (denoted $\prec$) for comparing two elements of $a, b \in X$ such that $a \prec b, b \prec a$ or $a = b$. For example, one could define a strict total ordering of the alphabet based on the existing order, i.e., $a \prec b \prec c \prec \cdots \prec z$.

However, the more general definition of well-ordering is not required in this book.**]**

While not obvious, it is well known (by mathematicians) that the well-ordering principle is equivalent to the previously mentioned axiom of choice, see the Wikipedia article on this topic [34].

Theorem 9-1 (Archimedean Property) For positive integers x and y, there exists a positive integer z such that $zx \geq y$.

Proof: Proof by contradiction is used here, i.e., assume the conclusion is false and derive a contraction (thus implying that the conclusion must be true).

So, assume that there exists a pair of positive integers x and y such that $zx < y$ for every positive integer z. Next, define the set $A = \{y - zx, \text{for } z = 1,2,3, \dots\}$. It follows (by the assumption $zx < y$) that A only contains positive integers. So, by the well-ordering principle, we know that A must have a least element (say $y - wx$). We also have that $y - (w+1)x \in A$ (by the definition of A). Further, $y - (w+1)x = (y - wx) - x < y - wx$ but this contradicts the choice of $y - wx$ as the least element in S. So, our initial assumption must be false and in fact, there must exist a positive integer z such that $\mathrm{zx} \geq \mathrm{y}$, which was to be proved■

9.1.2 Principle of Finite Induction

The principle of finite induction is very important in mathematics as it is used to prove many theorems. By way of analogy, finite induction is like dominoes. If you know (1) the dominoes are equally spaced so that if any given domino falls, the next will fall and so on, and (2) the first domino has fallen, then you can conclude eventually every domino will fall. In finite induction, we have a statement with variable n (rather than a domino). If the statement can be shown true for $n = 1$, and it can be proved that "if the statement is true for $n = k$ then the statement is true for $n = k + 1$," then finite induction tells us the statement is true for all values of n.

Another analog comes from the book Concrete Mathematics [35]

> Mathematical induction proves that we can climb as high as we like on a ladder, by proving that we can climb onto the bottom rung (the basis) and that from each rung we can climb up to the next one (the step).

Although its name may suggest otherwise, mathematical induction should not be considered as a form of inductive reasoning as defined earlier in this book. Mathematical induction is, in fact, an example of a deductive reasoning technique. The confusing terminology is unfortunate but the terms "inductive reasoning" and "mathematical induction" are firmly embedded in the literature and not likely to change.

The principle of finite induction is stated more formally in the following theorem:

Theorem 9-2 (First Principle of Finite Induction) Let S be a set of positive integers such that

- $1 \in S$
- *whenever* $k \in S$, *it must be that* $k + 1 \in S$

then S is necessarily the set of all positive integers.

Proof: By way of contradiction, assume that the set T (of all positive integers not in S) is nonempty. By the well-ordering principle, T must have a least element (call it x). We are given that $1 \in S$ and so it must be that $x > 1$ and thus, $0 < x - 1 < x$. Since x is the least element in T, $x - 1 \notin T$ which implies that $x - 1 \in S$. By hypothesis, S must contain $(x - 1) + 1 = x$ which contradicts the fact that $x \in T$. So, T must be empty and thus S is the set of all positive integers■

As an easy illustration of the first principle of finite induction, we prove that the sum of the first n odd numbers is n^2, i.e., $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ (for all positive integer values of n).

> **Proof**: Clearly, the formula holds for $n = 1$. Assume the formula is true for $n = k$, i.e., $1 + 3 + 5 + \cdots + 2k - 1 = k^2$. Consider the case of $n = k + 1$, i.e., $[1 + 3 + 5 + \cdots + (2k - 1)] + (2k + 1) = k^2 + (2k + 1) = (k + 1)^2$ (which was to be proved)■

There is an alternate version of the principle of finite induction that strengthens the second hypothesis:

Theorem 9-3 (Second Principle of Finite Induction) Let S be a set of positive integers such that

- $1 \in S$
- *whenever* $1,2, \ldots, k \in S$, *it must be that* $k + 1 \in S$

then S is necessarily the set of all positive integers.

Proof: By way of contradiction, assume that set T (of all positive integers not in S) is nonempty. By the well-ordering principle, T must have a least element (call it x). By hypothesis, x must be greater than 1. Further, since x is the least element in T, $1,2, \ldots, (x - 1)$ are not in T and are thus in S. But the second hypothesis implies that $(x - 1) + 1 = x \in S$ which contradicts the fact that $x \in T$. So, T must be empty and thus S is the set of all positive integers■

Some statements do require the second principle of finite induction (as opposed to the first principle of induction). For example, consider the Lucas sequence: 1,3,4,7,11,18,29,47,76, …

The general pattern (after the first two terms) is $x_n = x_{n-1} + x_{n-2}$ (basically add the previous two numbers to get the next number in the sequence). We use the second principle of finite induction

to prove that $x_n < \left(\frac{7}{4}\right)^n$. We have for $n = 1$ that $x_1 = 1 < \left(\frac{7}{4}\right)^1$. Next, assume that the statement holds for $n = 1,2, \ldots, k-1$. This gives us $x_{k-1} < \left(\frac{7}{4}\right)^{k-1}$ and $x_{k-2} < \left(\frac{7}{4}\right)^{k-2}$. It then follows that

$$x_k = x_{k-1} + x_{k-2} < \left(\frac{7}{4}\right)^{k-1} + \left(\frac{7}{4}\right)^{k-2} = \left(\frac{7}{4}\right)^{k-2}\left(\frac{7}{4}+1\right) = \left(\frac{7}{4}\right)^{k-2}\left(\frac{11}{4}\right) < \left(\frac{7}{4}\right)^{k-2}\left(\frac{7}{4}\right)^2 = \left(\frac{7}{4}\right)^k$$

Thus, given that the statement is true for $n = 1,2, \ldots, k-1$, we have proven the statement true for the case $n = k$, and the statement must therefore hold true for all values of n by the second principle of finite induction■

9.1.3 Binomial Theorem

The binomial theorem describes the expansion of powers of the sum of two variables, i.e., $(x+y)^n$. The binomial theorem arises in many places, including several other sections of this book, i.e., the sections on combinatorics and probability.

Before we state and prove the binomial theorem, we need some definitions.

The **factorial** of a positive integer n (written as $n!$) is the product of all positive integers up to and including that number, i.e., $n! = 1 \cdot 2 \cdot 3 \ldots \cdot n$.

When $(x+y)^n$ is expanded, the coefficients of the various terms are called **binomial coefficients**. For example, if we expand $(x+y)^3$ to get $x^3 + 3x^2y + 3y^2x + y^3$, then the binomial coefficients are 1, 3, 3 and 1. As another example, $(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4y^3x + y^4$ and so the binomial coefficients in this case are 1, 4, 6, 4, and 1. If one lists the binomial coefficient for successive powers of $(x+y)$, a pattern emerges, as shown in Figure 27. Notice that a given coefficient is the sum of the two coefficients directly above (to the left and right). As shown in bold in the figure, 10 is the sum of 4 and 6. This pattern is true throughout. The pattern in Figure 27 is known as Pascal's triangle.

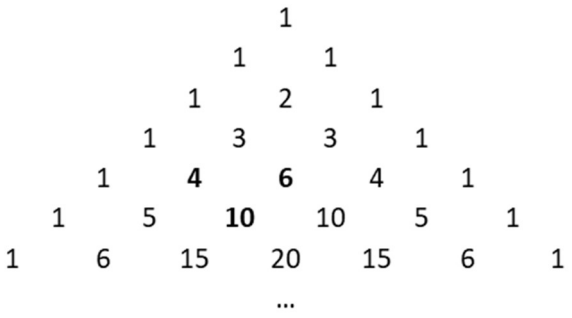


Figure 27. Pascal's Triangle

The pattern can be expressed in terms of factorials. Entry $k+1$ in row $n+1$ is given by

$$\frac{n!}{(n-k)!\,k!}$$

which is written in the shorthand notation as $\binom{n}{k}$. For example, to find the 3rd coefficient in the 6th row, one computes $\binom{5}{2} = \frac{5!}{(2!)(3!)} = 10$.

Further, the pattern concerning an entry being the sum of the two entries above it (to the left and right) in the above row can be written as an equation, i.e.,

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

This is known as **Pascal's rule**, which is easy to prove via algebra:

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{(n-k-1)!\,k!} + \frac{(n-1)!}{(n-k)!\,(k-1)!}$$
$$= \frac{(n-1)!\,(n-k)}{(n-k)!\,k!} + \frac{(n-1)!\,k}{(n-k)!\,k!}$$
$$= \frac{(n-1)!}{(n-k)!\,k!}(n-k+k) = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}$$

Now we are in a position to prove the following theorem.

Theorem 9-4 (Binomial Theorem)

$$(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n$$

or this can be written using the summation notation as

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

Proof: Proof is by the finite induction.

For the case $n = 1$, we have

$$(x+y)^1 = \sum_{k=0}^{1}\binom{1}{k}x^{1-k}y^k = \binom{1}{0}x^1y^0 + \binom{1}{1}x^0y^1 = x + y$$

As per induction, assume the formula holds for $n = m$ and prove that this implies the formula must hold true for $n = m + 1$.

Notice that $(x+y)^{m+1} = x(x+y)^m + y(x+y)^m$. Since we assumed the formula holds for $n = m$, we have the following two equations:

$$x(x+y)^m = \sum_{k=0}^{m}\binom{m}{k}x^{m-k+1}y^k = x^{m+1} + \sum_{k=1}^{m}\binom{m}{k}x^{m-k+1}y^k$$
$$y(x+y)^m = \sum_{j=0}^{m}\binom{m}{j}x^{m-j}y^{j+1} = \sum_{k=1}^{m}\binom{m}{k-1}x^{m-k+1}y^k + y^{m+1}$$

In the second equation above, we did a bit of a trick in adjusting the index to go from 1 to m (basically substituting $k-1$ for j.

Adding the above two equations and using Pascal's rule gives

$$(x+y)^{m+1} = x^{m+1} + \sum_{k=1}^{m}\left[\binom{m}{k} + \binom{m}{k-1}\right]x^{m-k+1}y^k + y^{m+1} = \sum_{k=0}^{m+1}\binom{m+1}{k}x^{m-k+1}y^k$$

We have shown the formula is true for the case $n = m+1$ given that it is true for the case $n = m$ and thus, by the principle of finite induction, the formula is true for all positive integer values of n■

9.2 Divisibility

In number theory, the term "**divisibility**" means that one integer exactly divides another (with no remainder). If x exactly divides y, we write $x|y$. This means that $\exists z \in \mathbb{Z} \ni y = zx$ or in words, there exists some integer z such that $y = zx$. If x does not exactly divide y, then we write $x \nmid y$.

Several basic facts can be proved concerning divisibility, as noted in the following theorem.

Theorem 9-5 For integers a, b and c, the following statements hold true:

a. $a|0, 1|a, a|a$

b. $a|1 \Leftrightarrow a = \pm 1$

c. $(a|b \text{ and } c|d) \Rightarrow ac|bd$

d. $a|b \text{ and } b|c \Rightarrow a|c$ (transitivity)

e. $(a|b \text{ and } b|a) \Leftrightarrow a = \pm b$

f. $(a|b \text{ and } b \neq 0) \Rightarrow |a| \leq |b|$

g. $(a|b \text{ and } a|c) \Rightarrow a|(bx+cy) for\ any\ integers\ x\ and\ y$

Proof: We prove Properties d and g, leaving the others as exercises for the reader.

(Transitivity) We are given that $a|b$ which implies there exists an integer m such that $b = am$. Similarly, $b|c$ implies there exists an integer n such that $c = bn$. So, we have $c = bn = a(mn)$ and thus $a|c$ (by definition of divisibility).

To prove Property g, note that $a|b \Rightarrow \exists m \in \mathbb{Z} \ni b = am$ and $a|c \Rightarrow \exists n \in \mathbb{Z} \ni c = an$. For any integers x and y, we can write $bx = amx$ and $cy = any$ which implies $bx + cy = amx + any = a(mx + ny)$ and by the definition of divisibility $a|bx + cy$, which was to be proved■

As we know from elementary arithmetic, when a number does not exactly divide another, a remainder is left. This fact is formalized in the following theorem.

Theorem 9-6 (Division Algorithm) Given integers a and b, with $b \neq 0$, there exist unique integers q (the quotient) and r (remainder) such that $a = qb + r$ and $0 \leq r < |b|$.

9.2.1 Greatest Common Divisor (GCD)

The **greatest common divisor** of two integers a and b, written as $\gcd(a, b)$, is the largest number that exactly divides both a and b. For example, $gcd(4,6) = 2, gcd(7,11) = 1, gcd(6,35) = 1$ and $gcd(-4,-16) = 4$. Several subtle points are worth mentioning:

- Two numbers do not need to be prime (no divisors other than 1) in order to have a gcd of 1, e.g., 6 and 35.

- The gcd is always a positive number. So, even though, -4 and -16 are both negative, their greatest common divisor is 4 since 4 does, in fact, exactly divide both numbers.

Theorem 9-7 (Bézout's identity) For any integers a and b (both not zero), there exists integers x and y such that $gcd(a,b) = ax + by$.

Proof: Define the set $A = \{aw + bz: aw + bz > 0 \text{ and } w, z \; are \; integers\}$. Assume, without loss of generality (wlog), that $a \neq 0$. Then $aw + b \cdot 0 \in A$ where w is set to -1 if $a < 0$ and to 1 if $a > 0$. So, $a \in A$ or $-a \in A$ and thus A is not empty. By the well-ordering principle, A has a smallest element (call it g). Since $g \in A$, there exists x and y such that $g = ax + by$. We next show that $g = \gcd(a,b)$.

From the division algorithm (Theorem 9-6), we know there exists integers q and r such that $a = qg + r$ with $0 \leq r < g$. So, $r = a - qg = a - q(ax + by) = a(1 - qx) + b(-qy)$.

If $r > 0$, then $r \in A$ which contradicts g being the least element of A. So, it must be that $r = 0 \Rightarrow a = qg \Rightarrow g|a$. Similarly, we can show that $g|b$.

By part (g) of Theorem 9-5, if h is any positive divisor of a and b, $h|(ax + by) = g$. So, by part (f) of Theorem 9-5 $h = |h| \leq |g| = g$. Thus, we have shown that $g = \gcd(a,b)$ ■

The solution (in terms of x and y) in Theorem 9-7 is not unique. For example, if $a = 15$ and $b = 5$, the $\gcd(15,5) = 5$ and we have, for example, $1 \cdot 15 - 2 \cdot 5 = 2 \cdot 15 - 5 \cdot 5 = -1 \cdot 15 + 4 \cdot 5 = 5$.

In some cases, the only common factors of two integers may be -1 and 1. In such cases, the two integers are said to be **relatively prime** and their gcd is 1. For example, $gcd(16,25) = 1$. It should be emphasized the neither 16 nor 25 are prime numbers. Of course, the gcd of two primes numbers is necessarily 1.

The following theorem gives an alternate criterion for two (non-zero) integers to be relatively prime in terms of a linear combination of the two numbers.

Theorem 9-8 Non-zero integers a and b are relatively prime if and only if there exists integers x and y such that $ax + by = 1$.

Proof: Going in one direction, if $\gcd(a,b) = 1$ then we have from Theorem 9-7 that there exists x and y such that $ax + by = 1$.

Going in the other direction, assume there exists x and y such that $ax + by = 1$. Further, let $\gcd(a,b) = g$ which implies $g|a$ and $g|b$. By part (g) of Theorem 9-5, $g|ax + by = 1$ and since g is a positive integer (as are all $gcds$), $g = 1$■

If two integers each divide a third integer (i.e., $a|c$ and $b|c$), does the product also divide the third (i.e., $ab|c$)? In general, the answer is no. For example, $6|12$ and $4|12$ but $4 \cdot 6 = 24 \nmid 12$. The required condition is that a and b need to be relatively prime, see Theorem 9-10.

The following two theorems (which rely on Theorem 9-8) will come in handy concerning the proofs of some other theorems.

Theorem 9-9 If $gcd(a,b) = g$, then $gcd\left(\frac{a}{g}, \frac{b}{g}\right) = 1$.

Proof: Note that $\frac{a}{g}$ and $\frac{b}{g}$ are whole numbers since by definition, $g|a$ and $g|b$.

From Theorem 9-7, we have that there exist integers x and y such $g = ax + by$ which can be written as $1 = \left(\frac{a}{g}\right)x + \left(\frac{b}{g}\right)y$. From Theorem 9-8, we have that $\frac{a}{g}$ and $\frac{b}{g}$ are relatively prime, i.e., their gcd is 1■

Theorem 9-10 If $a|c, b|c$ and $gcd\ (a,b) = 1$, then $ab|c$.

Proof: By definition of divisibility, there exists m and n such that $c = an$ and $c = mb$. By Theorem 9-8, there exists integers x and y such that $ax + by = 1$. Multiplying the previous equation by c and then substituting $c = an$ and $c = mb$, we get

$$c = acx + bcy = a(mb)x + b(an)y = ab(mx + ny)$$

which implies $ab|c$■

If an integer divides a product ($a|bc$), then it is not necessarily true that $a|b$ or $a|c$. For example, $15|30 = 3 \cdot 10$ but $15 \nmid 3$ and $15 \nmid 10$. As described in the following theorem, the required condition is that a and b be relatively prime. (The theorem is stated compactly using notation that we learned previous in the section on propositional logic.)

Theorem 9-11 (Euclid's lemma) $[(a|bc) \wedge (gcd(a,b) = 1)] \Rightarrow a|c$

Proof: By Theorem 9-8, there exists integers x and y such that $ax + by = 1$. Multiply the equation by c to get $c = (ax + by)c = acx + bcy$. Since $a|ac$ and $a|bc$ (given in the statement of the theorem), we have that $a|(acx + bcy) = c$■

9.2.2 Euclid's Algorithm

While it is possible to determine the gcd of two smaller numbers by trial and error, some sort of algorithm is needed for larger numbers. Such an algorithm does exist – it is known as Euclid's algorithm. The approach is based on the quotient algorithm (Theorem 9-6) and Theorem 9-12 (which follows below).

Consider the problem of finding the $\gcd(a,b)$. Apply the quotient algorithm to a and b, and to successive pairs of remainders as follows:

$$a = q_1 b + r_1, 0 < r_1 < b$$
$$b = q_2 r_1 + r_2, 0 < r_2 < r_1$$
$$r_1 = q_3 r_2 + r_3, 0 < r_3 < r_2$$
$$\ldots$$
$$r_{n-2} = q_n r_{n-1} + r_n, 0 < r_n < r_{n-1}$$
$$r_{n-1} = q_n r_n + 0$$

Notice that the sequence $r_1, r_2, r_3, \ldots, r_{n-1}, r_n$ is strictly decreasing and will eventually converge to 0. Further, the following theorem implies that the $gcd(a,b) = gcd(b,r_1) = \gcd(r_1,r_2) = \cdots = \gcd(r_{n-1},r_n) = r_n$.

Theorem 9-12 If for integers a, q, b and r, we have $a = qb + r$, then $gcd(a, b) = gcd(b, r)$.

Proof: If we let $g = \gcd(a, b)$, then $g|a$ and $g|b$. From Theorem 9-5 Part g, we have that $g|(1 \cdot a - qb) = r$. Thus, $g|b$ and $g|r$. We just need to show that g is the greatest divisor of b and r. To that end, take any other divisor of b and r (say h). By Theorem 9-5 Part g, we have that h divides any linear combination of b and r, and in particular, $h|(qb + r) = a$. Since $h|a, h|b$ and $g = \gcd(a, b)$, it must be that $h \leq g$ and thus $g = \gcd(b, r)$ ■

The algorithm should be clearer when used with a specific example. Consider the problem of finding $\gcd(88352,12364)$ using Euclid's algorithm:

$$88352 = 7 \cdot 12364 + 1804$$
$$12364 = 6 \cdot 1804 + 1540$$
$$1804 = 1 \cdot 1540 + 264$$
$$1540 = 5 \cdot 264 + 220$$
$$264 = 1 \cdot 220 + 44$$
$$220 = 5 \cdot 44$$

So, $gcd(88352, 12364) = gcd(12364,1802) = \cdots = gcd(220,44) = 44$.

Euclid's algorithm can be used to prove the following theorem, which is helpful when manually determining the gcd of two numbers that have some obvious common factors (e.g., if both numbers are even).

Theorem 9-13 $gcd(ka, kb) = k \cdot gcd(a, b)$, for $k > 0$.

Proof: In the derivation of Euclid's algorithm for the gcd of a and b, multiple each line by k to get

$$ak = q_1(bk) + r_1k, 0 < r_1k < kb$$
$$bk = q_2(r_1k) + r_2k, 0 < r_2k < r_1k$$
$$r_1k = q_3(r_2k) + r_3k, 0 < r_3k < r_2k$$
$$\ldots$$
$$r_{n-2}k = q_n(r_{n-1}k) + r_nk, 0 < r_nk < r_{n-1}k$$
$$r_{n-1}k = q_n(r_nk) + 0$$

However, the above sequence of steps is the application of Euclid's algorithm to the integers ak and bk, and so we have $gcd(ka, kb) = r_nk = k \cdot \gcd(a, b)$ ■

For example, in the previous problem concerning $\gcd(88352,12364)$, we could have simplified the problem by factoring out the 4 and 11 to get $gcd(88352, 12364) = 44 \cdot \gcd(2008, 281)$. Application of Euclid's algorithm to 2008 and 281 reveals that the two numbers are relatively prime. In Section 9.2.4, a simple test for divisibility by 11 is provided which can be used to simplify some gcd problems.

9.2.3 Least Common Multiple (LCM)

The **least common multiple** of two integers a and b, written as $lcm(a, b)$, is the smallest positive integer that is divisible by both a and b. For example, $lcm(2,7) = 14, lcm(3,6) = 6, lcm(6,8) = 24$ and $lcm(15, -20) = 60$. In general, $lcm(a, b) \leq |ab|$.

The gcd and lcm are related in a very simple way, as described in the following theorem.

Theorem 9-14 For positive integers a and b, $gcd(a,b) \cdot lcm(a,b) = ab$.

Proof: Let gcd$(a,b) = d$. By Theorem 9-9, $e = \frac{a}{d}$ and $f = \frac{b}{d}$ are relatively prime. So, we have $a = ed$ and $b = fd$. If we can show that $lcm(a,b) = def = d\left(\frac{a}{d}\right)\left(\frac{b}{d}\right) = \frac{ab}{d} = ab\backslash gcd(a,b)$, then we will be done.

Since $def = af$ and $def = be$, def is a common multiple of a and b.

Next, let s be any common multiple of a and b. If we can show that $def|s$, then def must be the $lcm(a,b)$.

Since s is a multiple of a, we have $s = ka = ked$ for some integer k (Equation 1).

We also have that $b|s$, and so $b = fd|s = ked$, which implies $fd|ked$. Cancelling the d from both sides of $fd|ked$ we get $f|ke$. So, we have that $f|ke$ and that f and e are relative prime. Thus, by Theorem 9-11, $f|k$ which (by definition) implies there exist some integer n such that $k = fn$. Now, substitute $k = fn$ into Equation 1 to get $s = fned = n(def)$ equivalently, $def|s$ ■

9.2.4 Divisibility Tests

Various simple tests are available for the quick determination of whether a given number (even a large number) is divisible by a single digit or in some cases a two or three digit number. In what follows, an n-digit number will be represented as

$$a = a_{n-1}a_{n-2}\ldots a_1a_0 = a_{n-1}\cdot 10^{n-1} + a_{n-2}\cdot 10^{n-2} + \cdots + a_1\cdot 10^1 + a_0$$

Divisibility by 10: If $a_0 = 0$, then clearly the number a is divisible by 10.

Going in the other direction, assume $10|a$. We also have that $10|(a_{n-1}\cdot 10^{n-1} + a_{n-2}\cdot 10^{n-2} + \cdots + a_1\cdot 10^1)$ and so, by Theorem 9-5 (g), we have that $10|[a - (a_{n-1}\cdot 10^{n-1} + a_{n-2}\cdot 10^{n-2} + \cdots + a_1\cdot 10^1)] = a_0$.

Thus, a number is divisible by 10 if and only if its least significant digit is a zero.

Divisibility by 3: Assume the sum of the digits of a is divisible by three, i.e., $3|(a_{n-1} + a_{n-2} + \cdots + a_1 + a_0)$. Rewrite the integer a as follows:

$$[a_{n-1}\cdot(10^{n-1} - 1) + a_{n-2}\cdot(10^{n-2} - 1) + \cdots + a_1\cdot(10^1 - 1)] + (a_{n-1} + a_{n-2} + \cdots + a_1 + a_0).$$

Further, $10^k - 1 = 99\ldots9$ (k nines) which is divisible by three. So, the left-hand side of the above expression is divisible by 3 and we are given that the right-hand side of the equation is divisible by 3. Thus, $3|a$.

This is true in the other direction. If we are given $3|a$ and we already know $3|[a_{n-1}\cdot(10^{n-1} - 1) + a_{n-2}\cdot(10^{n-2} - 1) + \cdots + a_1\cdot(10^1 - 1)]$, then we have that 3 must divide the difference of the above term with a, which is $a_{n-1} + a_{n-2} + \cdots + a_1 + a_0$.

Divisibility by 11: Assume the alternating sum of the digits of a is divisible by 11, i.e., $11|(-1)^n a_{n-1} \pm \cdots - a_3 + a_2 - a_1 + a_0$. To be clear, we start from the right, with the least significant digit a_0 always being positive, and then alternating from minus to plus for higher order terms. For example, consider 45938 and the alternating sum of its digits, i.e., $4 - 5 + 9 - 3 + 8 =$

13. Thus $11 \nmid 45938$. However, for 19293835, we have $-1+9-2+9-3+8-3+5=22$ and so, $11|19293835$. In fact, $19293835 = 11 \cdot 1753985$.

Notice that 11 evenly divides $10+1, 10^2-1, 10^3+1, 10^4-1$ and so on. In general, we have that $11|(10^k-(-1)^k)$. This is true since $10 \equiv -1(mod\ 11)$ and by Theorem 6-10 Part c, we get the result $10^k = (-1)^k(mod\ 11)$ which is equivalent to $11|(10^k-(-1)^k)$.

This leads us to rewrite a as

$[a_{n-1} \cdot (10^{n-1}-(-1)^{n-1}) + \cdots + a_3(10^3+1) + a_2(10^2-1) + a_1 \cdot (10^1+1)] + (\pm a_{n-1} + \cdots - a_3 + a_2 - a_1 + a_0)$.

The left-hand side of the above expression (in between the square brackets) is divisible by 11 and we were given that the right-side is divisible by 11, and thus $11|a$.

For an extensive listing of divisibility tests, see the Wikipedia article on Divisibility [36].

9.3 Diophantine Equations

Consider the following ancient number theory puzzles (all with similar solutions):

- (Attributed to Alcuin of York, 775 AD) One hundred bushels of grain are distributed to 100 people in three groups (A, B and C) such that each person in Group A receives 3 bushels, each person in Group B receives 2 bushels and each person in Group C receives ½ bushel. How many people are in each group? If we let x, y and z be the number of people in Groups A, B and C, respectively, then we have $x+y+z=100$ and $3x+2y+\frac{1}{2}z=100$.
- (Attributed to Mahaviracarya, 850 AD) There are 63 baskets of bananas (each with the same number of bananas) and 7 additional bananas. The bananas are distributed equally among 23 people. How many bananas are there in each basket? Let x equal the number of bananas in each basket and y equal the number of bananas distributed to each person, then $63x+7=23y$.
- (Attributed to Yen Kung, 1372 AD) Given a collection of coins (of unknown number). If 77 equal stacks of the coins are made, 27 coins are left over. However, if 78 equal stacks are made, there are no coins left over. How many coins are there? If we let x equal the number of coins in the first type of stack and y equal the number of coins in the second type of stack, then the total number of coins equals $77x+27=78y$.

With some rearrangement and simplification, all of the above problems can be reduced to solving an equation of the form $ax+by=c$ where a, b and c are integers. This is known as a Diophantine equation, named after the Hellenistic mathematician of the 3rd century AD, Diophantus of Alexandria.

Theorem 9-15 The linear Diophantine equation $ax+by=c$ has an integer solution if and only if $gcd(a,b)\,|c$.

Proof: Let $g=gcd(a,b)$. This implies that $g|a$ and $g|c$, or equivalently, there exists integers r and s such that $a=gr$ and $b=gs$.

If a solution does exist to the equation, say $ax_0+by_0=c$, then we have

$$c = ax_0+by_0 = grx_0+gsy_0 = g(rx_0+sy_0)$$

which implies $g|c$.

Going in the other direction, assume that $g = gcd(a,b)\,|c$ or equivalently, $c = gt$ for some integer t. By Theorem 9-7, there exists integers x_0 and y_0 such that $g = ax_0 + by_0$. Multiply by t to get

$$c = gt = (ax_0 + by_0)t = a(tx_0) + b(ty_0)$$

which implies that $ax + by = c$ has $x = tx_0$ and $y = ty_0$ as a particular solution■

Further, once one solution is found to a linear Diophantine equation, it is possible to generate an infinite number of other solutions.

Theorem 9-16 If x_0, y_0 is a solution of the Diophantine equation $ax + by = c$ where $g = gcd(a,b)$ then all other solutions are given by

$$x = x_0 + \left(\frac{b}{g}\right)t,\ y = y_0 - \left(\frac{a}{g}\right)t,\ \textit{for any integer t.}$$

Proof: Assume there exists another solution to the equation (say x_1, y_1). Then we have

$$ax_0 + by_0 = c = ax_1 + by_1$$

which implies $a(x_1 - x_0) = b(y_0 - y_1)$ (Equation 1)

By Theorem 9-9, $\gcd\left(\frac{a}{g}, \frac{b}{g}\right) = 1$. Let $r = \frac{a}{g} \Rightarrow a = gr$ and let $s = \frac{b}{g} \Rightarrow b = gs$. Substituting into Equation 1 and cancelling g on both sides, we get

$$r(x_1 - x_0) = s(y_0 - y_1) \quad \text{(Equation 2)}$$

So, $r|s(y_0 - y_1)$ and $\gcd(r,s) = 1$. By Euclid's lemma, we have that $r|(y_0 - y_1)$ or equivalently, there exist an integer t such that $y_0 - y_1 = rt$. Substituting into Equation 2, we get $x_1 - x_0 = st$. Thus, we have

$$x_1 = x_0 + st = x_0 + \left(\frac{b}{g}\right)t$$

$$y_1 = y_0 - rt = y_0 - \left(\frac{a}{g}\right)t$$

So, if x_1, y_1 is a solution other than the given solution x_0, y_0 then it must be of the form given in the above equations.

We can verify that x_1, y_1 is a solution to the given Diophantine equation, regardless of the value of the integer t, by the following line of reasoning

$$ax_1 + by_1 = a\left[x_0 + \left(\frac{b}{g}\right)t\right] + b\left[y_0 - \left(\frac{a}{g}\right)t\right] = (ax_0 + by_0) + \left(\frac{ab}{g} - \frac{ab}{g}\right)t = c.$$

Thus, each integer value of t gives a solution to the equation■

We now have sufficient background to solve the problems stated at the beginning of this section. Let's take the baskets of bananas problem first. We can rewrite the related equation as $63x - 23y = -7$. Noting that 23 is a prime number, we know that $\gcd(63, -23) = 1$ which divides -7, and thus we know a solution exists to the problem. In order to determine a solution for the equation, we use the Euclidean algorithm:

$$63 = -2(-23) + 17$$
$$-23 = -2(17) + 11$$
$$17 = 1(11) + 6$$
$$11 = 1(6) + 5$$
$$6 = 1(5) + 1$$
$$5 = 5(1)$$

This tells us that $\gcd(63, -23) = 1$ which we already knew, **but** we can also use the above calculations in reverse to determine a solution to $63x - 23y = -7$.

$$6 - 5 = 1$$
$$6 - (11 - 6) = 2(6) - 11 = 1$$
$$2(17 - 11) - 11 = 2(17) - 3(11) = 1$$
$$2(17) - 3(-23 + 2(17)) = -4(\mathbf{17}) + 3(23) = 1$$
$$-4\big(\mathbf{63} + \mathbf{2}(-\mathbf{23})\big) + 3(23) = -4(63) + 11(23) = 1$$
$$-4(63) - 11(-23) = 1$$

Multiply the last equation above by -7 to get $28(63) + 77(-23) = -7$ and thus, $x_0 = 28, y_0 = 77$ is a solution to the Diophantine equation. All solutions are given by

$$x = 28 - 23t$$
$$y = 77 - 63t$$

The smallest positive solution is obtained when $t = 1$, i.e., $x = 5, y = 14$. So, there are 5 bananas in each of 63 baskets plus 7 more bananas, and each person gets 14 bananas.

Next, consider the bushels of grain problem attributed to Alcuin of York. We have that $x + y + z = 100$ which we can rewrite as $z = 100 - x - y$ and then substitute into $3x + 2y + \frac{1}{2}z = 100$ to get $3x + 2y + \frac{1}{2}(100 - x - y) = 100$ which simplifies to $5x + 3y = 100$. This equation is simple enough to guess at a solution, i.e., $x = 5, y = 25$, which implies $z = 70$. There are no other solutions that meet the additional requirement that $x + y + z = 100$.

The problem attributed to Yen Kung is left as an exercise for the reader.

9.4 Prime Numbers

A **prime number** is a positive integer that is only divisible by 1 and itself. The number 1 is not considered to be prime. Non-prime positive integers are called **composite numbers**. For example, all the even numbers (except for 2) are composite. The first few prime numbers are 2, 3, 5, 7, 11, 13, 17, 19.

The proof of the next theorem provides a good example of the use of the second principle of finite induction.

Theorem 9-17 If p is a prime and $p|a_1 \cdot a_2 \cdot \ldots \cdot a_n$ then $p|a_k$ for some k, where $1 \leq k \leq n$.

Proof: For $n = 1$, the statement is trivially true, i.e., $p|a_1$.

Assume the statement is true for $n = 1, 2, \ldots k-1$ and prove that it holds for $n = k$, then by induction, the statement holds for all values of n.

So, assume $p|a_1 \cdot a_2 \cdot \ldots \cdot a_k$. Let $b = a_1 \cdot a_2$, then we have that $p|b \cdot a_3 \cdot a_4 \cdot \ldots \cdot a_k$ but this expression has only $k-1$ terms. So, by the induction hypothesis, p divides one of the terms in the set $\{b, a_3, a_4, \ldots, a_k\}$. If $p|a_j, j = 3,4, \ldots, k$ then we are done. Otherwise, $p|b = a_1 \cdot a_2$, but the induction hypothesis holds for $n = 1, 2, \ldots, k-1$. Thus, $p|a_1$ or $p|a_2$. In any event, $p|a_j$ for some $j \in \{1,2, \ldots, k\}$ and so, the theorem is proved■

We now state one of the most important and basic theorems of number theory.

Theorem 9-18 (Fundamental Theorem of Arithmetic) Every positive integer $n > 1$, can be expressed as a product of primes and the representation is unique (except for the order in which the factors appear). In particular, one can represent n in the following canonical form:

$n = p_1^{k_1} \cdot p_2^{k_2} \ldots p_r^{k_r}$ where $p_1 < p_2 < \cdots < p_r$ and $p_1, p_2, \ldots, p_r$ are prime numbers.

For example, $14700 = 2^2 \cdot 3 \cdot 5^2 \cdot 7^2$.

As an immediate application of the above theorem, we prove that if s is a prime number, then $\sqrt{s}$ is an irrational number (i.e., cannot be expressed as a fraction).

Theorem 9-19 $\sqrt{s}$ is an irrational number where s is a prime number.

Proof: By way of contradiction, assume that $\sqrt{s}$ is rational, i.e., there exists integers r and q such that $\sqrt{s} = \frac{r}{q}$. Without loss of generality (wlog), assume that $\frac{r}{q}$ is reduced such that $\gcd(r, q) = 1$. Squaring both sides of $\sqrt{s} = \frac{r}{q}$ and the multiplying both sides by q^2 gives $sq^2 = r^2$. This means that r^2 is a multiple of s, i.e., $s|r^2$. By the fundamental theorem of arithmetic, r can be uniquely expressed as a product of prime numbers, i.e., $r = p_1^{k_1} \cdot p_2^{k_2} \ldots p_r^{k_r}$. Thus, $r^2 = p_1^{2k_1} \cdot p_2^{2k_2} \ldots p_t^{2k_t}$. Since s is prime and divides r^2, Theorem 9-17 tells us that s must be one of the elements in $\{p_1, p_2, \ldots, p_t\}$ but this means that $s|r$ which implies there an integer m such that $r = ms$. Substituting into $sq^2 = r^2$, we get $sq^2 = m^2s^2 \implies q^2 = m^2s \implies s|q^2$ and by a similar argument to the above, we get that $s|q$. So, $s|r$ and $s|q$ which contradicts the assumption $\gcd(r, q) = 1$. So, we've arrived at a contradiction based our assumption that $\sqrt{s}$ is rational and thus $\sqrt{s}$ must be irrational■

The fundamental theorem of arithmetic provides a way of calculating the gcd of two numbers. For example, consider $7425 = 3^2 \cdot 5^2 \cdot 11$ and $16752463 = 7^3 \cdot 13^2 \cdot 17^2$. We can easily see the $\gcd(7425, 16752463) = 1$. More revealing is the problem of finding the gcd of say 16752463 and $31111717 = 7^2 \cdot 13^3 \cdot 17^2$. The solution is gotten by taking the smallest power of each prime common to both terms, i.e., $gcd(16752463, 31111717) = 7^2 \cdot 13^2 \cdot 17^2 = 2393209$. This also works if the two numbers only share some prime factors in common, e.g.,

$$\gcd(5^6 \cdot 7^7 \cdot 13^2 \cdot 17^2, 7^2 \cdot 13^{10}) = 7^2 \cdot 13^2.$$

In what follows, we use the notation $\prod_{i \in A} X_i$ to mean "multiply the X_i terms together for all $i \in A$" where A is a set or some description of a property. For example, $\prod_{i \in \{1,2,3,4,5\}} i^2 = 1^2 \cdot 2^2 \cdot 3^2 \cdot 4^2 \cdot 5^2$.

Theorem 9-20 If the prime factorization of two numbers a and b are $a = \prod_i {p_i}^{a_i}$ (where i ranges over all the prime factors of a) and $b = \prod_j {p_j}^{b_j}$ (where j ranges over all the prime factors of b) then $gcd(a,b) = \prod_k {p_k}^{min(a_k,b_k)}$ (where k ranges over all the common prime factors of a and b).

This is a bit of an eye test, but it is just saying that to get the gcd of two numbers do the following:

- write the unique prime factorization of each number
- compare the common prime factors, and take the minimum power for each prime factor shared by both numbers and multiply these together.

We conclude this section with a theorem the first proof of which is attributed to Euclid. Even after 2000 years the proof stands as an excellent model of reasoning. See David Joyce's pages for an English translation of Euclid's actual proof .

Theorem 9-21 There are an infinite number of prime numbers.

Proof: Assume there are only a finite number of prime numbers, i.e., $p_1, p_2, \ldots, p_n$. Let

$$P = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$$

Since P is clearly larger than any of the assumed finite set of primes, it is not prime. Thus, P must be divisible by at least one of the primes (say p_r). When P is divided by p_r, the remainder is 1, implying $p_r \nmid P$. Thus, we have arrived at a contradiction and our original assumption must be false, and so, there must be an infinite number of prime numbers■

9.5 Exercises

1. Prove that $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. ***Hint**: Use the first principle of finite induction*.
2. Prove that $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$.
3. Use the first principle of finite induction to prove Part c of Theorem 6-10.
4. For prime numbers p and q, prove $p|a$ and $q|a \Leftrightarrow pq|a$. ***Answer**: Use Theorem 9-10 to prove in the forward direction.* Going in the other direction, assume $pq|a$ which is equivalent to an $a = pqn$ for some positive integer n. So, a can be written as $p(qn)$ which is equivalent to $p|a$. Similarly, a can be written as $q(pn)$ which is equivalent to $q|a$.
5. What is the minimum number of square towels that can be cut from a 24 by 40 piece of cloth without wasting any cloth? ***Hint**: Use the gcd function.* ***Answer:** Note that $\gcd(24,40) = 8$, then take $\frac{24(40)}{8^2} = 15$.*
6. At a given store, hot dogs come in packages of 6 and hot dog rolls come in packages of 8. What is the minimum number of packages of hot dogs and hot dog rolls that you need to purchase to have an exact match between hot dogs and rolls?
7. Use the binomial theorem to expand $(x + y)^5$.
8. Is the number 9294689855 divisible by 11?

9. Determine a test for divisibility by 5.
10. Prove that if the sum of the digits of a positive integer are divisible by 9, then the number is divisible by 9. ***Hint****: This is very similar to the derivation for divisibility by 3.*
11. Using reasoning similar that leading to Theorem 9-20, determine an expression for $lcm(a,b)$. Use your result to determine $lcm(x^7y^5z^9, w^3x^4y^8)$. ***Hint****: The answer is very similar to the gcd formula. Try some examples first and then reason to the general solution.*
12. Determine $gcd(x^7y^5z^9, w^3x^4y^8)$ where w, x, y and z are positive integers.
13. Use Euclid's algorithm to find $\gcd(10976,3969)$.
14. Solve the problem attributed to Yen Kung (as described in Section 9.3), i.e., find the minimal solution to $77x + 27 = 78y$ where x and y are positive integers. ***Answer****: All solutions are of the form $x = 27 - 78t$ and $y = 27 - 77t$. Minimum positive solution occurs when $t = 0$, $x = y = 27$.*
15. Find the first integer value for n such that $f(n) = n^2 + n + 1$ is a composite number. Do the same for $h(n) = 36n^2 - 810n + 2753$. *[**Remark****: It can be proven that there is no nonconstant polynomial $f(n)$ with integer coefficients that produces only prime number values for all integer values of n.]*

10 Combinatorics

10.1 Overview

Combinatorics is the branch of mathematics that entails the study of various methods for counting the number of things that fit a given set of conditions. For example, the determination of the number of possible 5-card hands from a deck of 52 playing cards is a combinatorial problem.

Combinatorics is essential for the computation of discrete probabilities. To compute the probability of an event (e.g., the probability of rolling an 11 with a pair of dice), one determines the number of all possible outcomes, the number of outcomes for the desired result and then divides the latter by the former. In the dice example, there are 36 possible outcomes but only 2 ways to roll an 11, and so the probability of rolling an 11 is $\frac{2}{36}$.

There is also a strong relationship between combinatorics and graph theory. In fact, much of combinatorics has been driven by graph theory problems involving counting.

10.2 Fundamentals

We start with several of the most fundamental rules of combinatorics.

Sum Rule for Counting: If an event (e.g., rolling a die) can happen in m ways and another event (e.g., drawing a card from a deck of 52 cards) can happen in n ways, and **only one** of the two events can happen at a given instance, then there are $m + n$ ways for one of two events to happen. For the example, there 58 possible outcomes for the die roll or card draw. This rule can be extended to r events of which **only one** can occur, where the i^{th} event can occur m_i ways. In this case, the total number of possible outcomes is $m_1 + m_2 + \cdots + m_r$.

In the case of the sum rule, there is an implied dependence between the events, i.e., if one event happens, the other events cannot occur.

Product Rule for Counting: If an event (e.g., rolling a die) can happen in m ways and another event (e.g., draw a card from a deck of 52 cards) can happen in n ways, and the two events are independent, then the number of possible outcomes for the two events is mn. Using the die roll and card drawing example again, but now assume both can happen simultaneously, we have a total of $6(52) = 312$ possible outcomes. This rule can be extended to r events, where the i^{th} event can occur in m_i ways. In this case, the total number of possible outcomes is $m_1 m_2 \ldots m_r$.

Pigeonhole Principle: If each item in a set of n items (call this Set #1) is to be associated with one item in another set (call this Set #2) where Set #2 has m items and $n > m$, then at least one item in Set #2 must have more than one item associated with it from Set #1. (Note that $m, n \in \mathbb{N}$.)

This seems terribly obvious (and it is), e.g., if you put 10 balls into (i.e., associate with) 9 boxes, then clearly one of the boxes must have 2 or more balls. The principle is nevertheless very useful, and it is not always so obvious as to when it can be applied.

Example 1: Suppose that you have a collection of many red marbles, blue marbles, green marbles and white marbles in a box. What is the least number of marbles that you need to retrieve from the box (while not looking) to be sure of getting two marbles of the same color?

Answer: At first glance, it is not clear how this problem relates to the pigeonhole principle. After some thought, one approach is to let the four colors be the pigeonholes. So, $m = 4$. We are being asked to find the smallest number n such that we are sure to draw two marbles of the same color. One can view Set #1 as the number of draws. The smallest value of n to ensure the desired result is 5. "Worst case" (assume that two marbles of the same color are desired), is to draw each a red, blue, green and white marble (in no particular order) in the first four draws. The next draw will force a match since there are only four colors.

Pigeonhole Principle – Generalization #1: If each item in a set of $k \cdot m + 1$ items (call this Set #1) is to be associated with one item in another set (call this Set #2) that has m items, then at least one item in Set #2 must have $k + 1$ items associated with it from Set #1. (Note that $k, m \in \mathbb{N}$.)

Example 2: Let Set #1 consist of $3(10) + 1 = 31$ balls (i.e., $k = 3, m = 10$) and Set #2 consist of 10 containers. The generalized pigeonhole principle tells us that at least one of the containers must have $k + 1 = 4$ balls. Figure 28 shows the best that one can do to avoid putting 4 balls into one container (given 31 balls). The 31st ball (at the top left of the figure) must go into one of the containers and thus, forcing 4 balls in one of the containers.

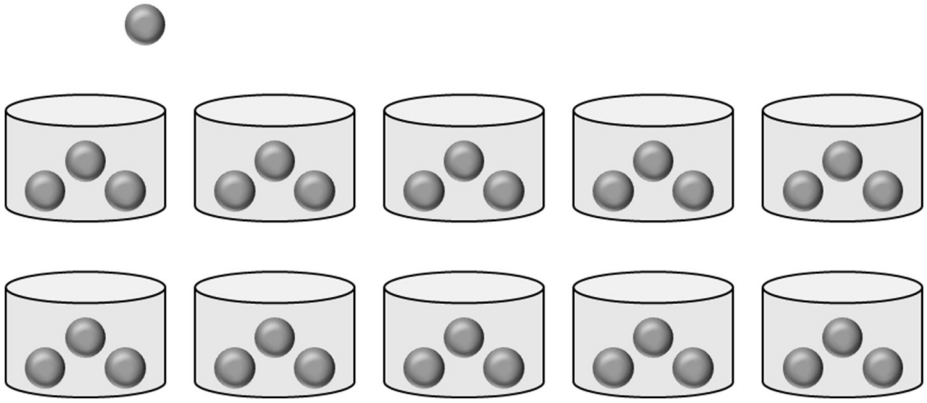


Figure 28. Pigeonhole Principle – Generalization #1 – Example 2

Example 3: Consider Example 1 regarding the box of marbles, but this time determine the least number of draws that are needed to ensure at least 5 marbles of the same color.

Answer: Using the Pigeonhole Principle – Generalization #1, we want $k + 1 = 5 \Rightarrow k = 4$. Further, as before, $m = 4$. So, we have $km + 1 = 17$, i.e., 17 draws are required to ensure at least 5 marbles of the same color. Think of it this way: if you selected only 16 marbles from the box, it is possible that you would get 4 of each color. The 17th selection forces 5 marbles of one of the colors. As shown in Figure 29, the containers represent colors and the balls represent selections of marbles of a given color. The best one can do without getting 5 marbles of the same color is by getting 4 of each color. The 17th draw will force at least 5 marbles of one color.

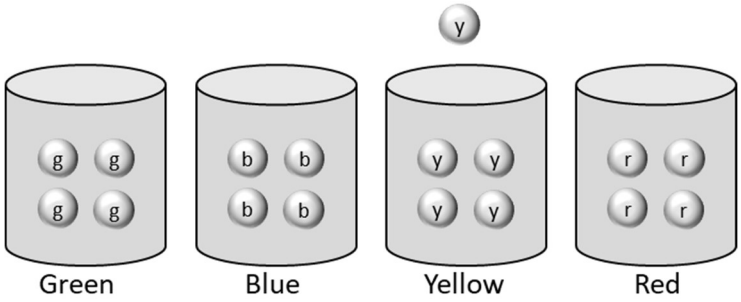


Figure 29. Pigeonhole Principle – Generalization #1 – Example 3

Pigeonhole Principle – Generalization #2: If there are $a_1 + a_2 + \cdots + a_n - n + 1$ or more items in a set (Set #1) to be associated with n items in another set (Set #2), then there exists a number $j \in \{1,2, \ldots n\}$ such that item j (in Set #2) has a_j or more associations with items in Set #1.

Think of the principle this way, i.e., you have n containers with capacities as shown in Figure 30. Assume that each container is filled to capacity for a total of $a_1 + a_2 + \cdots + a_n - n$ items. If you are forced to add one more item to any container, then at least one containers (say j) has a_j items.

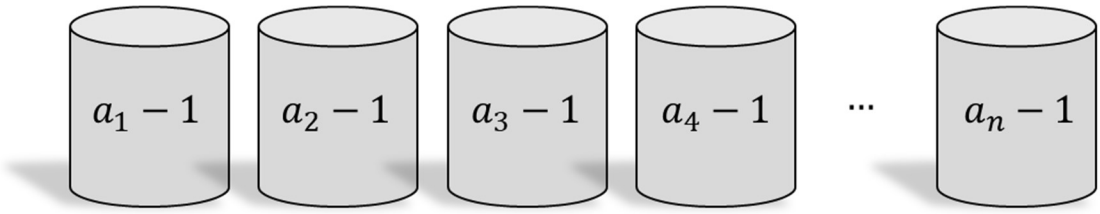


Figure 30. Pigeonhole Principle – Generalization #2

Proof: Assume to the contrary that $\forall j \in \{1,2, \ldots n\}$ item j (in Set #2) has at most $a_j - 1$ associations with items in Set #1. Summing the number of associations, we get at most $a_1 + a_2 + \cdots + a_n - n$ but this is one less than the number of items in Set #1, all of which were assumed to be associated with an item in Set #2. Thus, we have a contradiction and our contrary assumption is false. So, there must exist $j \in \{1,2, \ldots n\}$ such that item j (in Set #2) has a_j or more associations with items in Set #1■

Example 4: If someone wants their pocket change to have at least 4 half-dollars or at least 8 quarters or at least 20 dimes or at least 40 nickels, what is the least number of coins that this person needs to select from a jar of coins (assume there are more than 100 of each type of coin in the jar). Another way to state the problem is to ask: "what is the least number of coins that needs to be (blindly) selected from a jar of half-dollars, quarter, dimes and nickels to ensure $2 in change **of a particular coinage** (half-dollars, quarters, dimes or nickels)?"

> **Answer**: Using Pigeonhole Principle – Generalization #2, we have $n = 4$ and $a_1 = 4, a_2 = 8, a_3 = 20, a_4 = 40$. Thus, $a_1 + a_2 + a_3 + a_4 - n + 1 = 4 + 8 + 20 + 40 - 4 + 1 = 69$ is the minimum numbers of selections from the coin jar to ensure at least $2 in change of a particular coinage. For example (although improbably), one could draw 3 half-dollars, 7 quarters, 19 dimes and 39 nickels for a total of 68 coins and still not have $2 in any one coinage (failing one short in each case).

10.3 The Labeling Principle

10.3.1 Distinguishable Objects

Consider a set of 3 distinguishable objects (i.e., objects that are named or labeled in some way so that they can be distinguished). The objects can be anything, e.g., billiard balls, people, articles of clothing. The "labeling" can vary significantly, e.g., different shapes, different colors, different names – basically, any scheme that allows one object to be distinguished from another. For the sake of argument, we will just use letters (A, B and C) for our simple example. How many ways can A, B and C be arranged? For only 3 objects, it is easy to just list the possibilities, i.e., ABC, ACB, BAC, BCA, CAB and CBA.

The case for the arrangements of four distinguishable objects (call them A, B, C and D) reduces to the case of three. That is to say

- D comes first, with 6 ways to arrange A, B and C, or
- C comes first, with 6 ways to arrange A, B and D, or
- B comes first, with 6 ways to arrange A, C and D, or
- A comes first, with 6 ways to arrange B, C and D.

So, there are $4 \cdot 6 = 24$ ways to arrange 4 distinguishable objects.

For five distinguishable objects, we use the same logic as above to get $5 \cdot 4 \cdot 6 = 120$. Note that $6 = 3 \cdot 2 \cdot 1$, we can rewrite the answer for five distinguishable objects as $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$ which is abbreviated as $5!$ and read as "5 factorial." Recall that factorials were discussed earlier in the section on number theory regarding binomial coefficients.

In general, the number of arrangements for n distinguishable objects is given by

$$n! = n(n-1)(n-2) \ldots (3)(2)(1).$$

10.3.2 Indistinguishable Objects

If all the objects to be arranged are indistinguishable (e.g., AAAAA), there is only one possible solution.

The problem becomes more interesting when there is a mixture of distinguishable and indistinguishable objects. For example, how many arrangements are there of objects labelled as ABCCCD? One way to view the problem is to consider how many arrangements are possible if the Cs were different (say $C_1C_2C_3$) and then divide by the number of rearrangements lost given that the Cs are, in fact, indistinguishable. Consider the set of arrangements below where the positions of A, B and D are fixed, and temporarily assuming the Cs are distinguishable:

$$ABDC_1C_2C_3, ABDC_1C_3C_2, ABDC_2C_1C_3, ABDC_2C_3C_1, ABDC_3C_1C_2, ABDC_3C_2C_1$$

Arrangements of the three Cs for a given fixed arrangement (including position) of ABD leads to 3! arrangements, but this is true for each fixed arrangement of ABD. So, if the Cs are not distinguishable, then we need to divide the total number of possible arrangement (which is $6!$) by the number of identical arrangements because of the three identical letters (i.e., 3!). Thus, the answer is $\frac{6!}{3!}$.

Next, consider the more complex example with several repeated (identical) labels, e.g., AAAABBBCCD. We use the same logic as before but need to divide the total possible number of arrangements (for 10 distinct objects) by the number of arrangements lost because of the replicated terms. So, we get $\frac{10!}{(4!)(3!)(2!)(1!)}$. The 1! term is really not necessary (since it is just division by 1) but is a good bookkeeping practice (i.e., tracking of the D which is not repeated). Notice that the sum of the numbers in the denominator equals the numerator (ignoring the factorial sign).

The preceding discussion can be generalized and summarized in the following theorem:

Theorem 10-1 (Labeling Principle) Given n objects where a_1 of the objects have the label 1, a_2 of the objects have the label 2, … and a_r of the objects have the label r, then the total number of distinguishable arrangements is $\frac{n!}{(a_1!)(a_2!)\dots(a_r!)}$. Note that $n = a_1 + a_2 + \cdots + a_r$.

The general ideas in the preceding discussion about the labeling principle were developed by James Tanton, see Thinking Mathematics, Volume 2 [38]. The main idea is to simplify the sometimes confusing distinction between permutations and combinations. In fact, Tanton's approach avoids the concepts of permutation and combination.

[Author's Remark: James Tanton is a brilliant educator and expositor of mathematics. See his webpage (www.jamestanton.com) for a complete set of references to his books and videos.]

To reinforce the principle, consider the following examples.

Example 1: Given n distinguishable items, how many ways are there to select k of the items and by implication, not select $n - k$ of the items?

> **Answer**: Using the labeling principle, k items are labeled as "selected" and $n - k$ items are labeled as "not selected." This give a total of $\frac{n!}{(n-k)!k!}$ or just $\binom{n}{k}$, using the notation that we introduced earlier related to the binomial coefficient.

Example 2: Given a container with 10 lottery tickets, how many different ways can a 1st, 2nd and 3rd place winner be drawn?

> **Answer**: Using the labeling principle, 1 ticket is labeled as 1st, 1 ticket is labeled as 2nd place, 1 ticket is labeled as 3rd place and 7 tickets are labeled as "not winners". This gives a total of $\frac{10!}{(1!)(1!)(1!)(7!)} = 720$ possible results.

Example 3: Given a box with 100 different books, how many different ways are there to distribute 60 books to School #1 and 40 books to School #2?

> **Answer**: The books destined for School #1 are given the label 1 and the books destined for School #2 are given the label 2. From the labeling principle, there are $\frac{100!}{(60!)(40!)}$ to distribute the books. This is a huge number.

Example 4: Given 36 school children, how many ways can they be assigned to three softball teams (12 on each team) with one captain for Team #1, one captain for Team #2 and 2 co-captains for Team #3?

Answer: The labeling is as follows:

- 11 children labeled as Team #1, with 1 child labeled as Captain Team #1

- 11 children labeled as Team #2, with 1 child labeled as Captain Team #2
- 10 children labeled as Team #3, with 2 children labeled as Captain Team #3

From the labeling principle, there are $\frac{36!}{(11!)(1!)(11!)(1!)(10!)(2!)}$. This is also a huge number.

10.4 Problems Involving the Product Rule and Labeling Principle

Consider a deck of playing cards with 13 cards of each suit (clubs, spades, hearts and diamonds). How many ways are there to select 5 clubs, 6 spades, 7 hearts and 8 diamonds?

Answer: Selection of the clubs, spades, hearts are diamonds are independent events.

For the clubs, there are 5 labeled as "selected" and 8 as "unselected." From the labeling principle, there are $\frac{13!}{(8!)(5!)}$ ways to select the clubs.

Similarly, there are $\frac{13!}{(6!)(7!)}$ ways to select the spades, $\frac{13!}{(7!)(6!)}$ ways to select the hearts and $\frac{13!}{(5!)(8!)}$ ways to select the diamonds.

To get the final answer to the question, we use the product rule

$$\frac{13!}{(8!)(5!)}\frac{13!}{(6!)(7!)}\frac{13!}{(7!)(6!)}\frac{13!}{(5!)(8!)} = 1287^2 \cdot 1716^2$$

This is about 4.877 trillion.

How many ways are there to deal a full house (2 cards of the same value, and 3 cards of another value)? An example of a full house is Ace of Spades, Ace of Diamonds, 7 of Hearts, 7 of Spades and 7 of Clubs.

Answer: Label the card values as "S_1: selected for the 2-card part of a full house", "S_2: selected for the 3-card part of a full-house" and "N: not selected". From the labeling principle, we have $\frac{13!}{(1!)(1!)(11!)}$ ways of selecting the value for the 2-card and 3-card parts of the full house.

For the 2-card part of the full house, 2 cards are selected and 2 are not, giving $\frac{4!}{(2!)(2!)}$.

For the 3-card part of the full house, 3 cards are selected and 1 is not, giving $\frac{4!}{(3!)(1!)}$.

Using the product rule, we get the final solution:

$$\frac{13!}{(1!)(1!)(11!)}\frac{4!}{(2!)(2!)}\frac{4!}{(3!)(1!)} = 3744$$

How many ways are there to deal a "regular" straight (not necessarily of the same suit) from a deck of cards? Also, assume the Ace can play two roles, i.e., as a 1 or as the highest card in the deck above the King.

Answer: Label the card values as "selected as the first card value in the straight" or "not the first card value in the straight." Poker does not allow for wrapping around, i.e., the sequence Queen, King, Ace, Two, Three is not considered a straight. So, there are only 10 possible

choices for a starting card value in a straight. This is confirmed by the labeling principle, i.e., $\frac{10!}{(1!)(9!)} = 10$.

For each card value in the straight, one card is selected (a particular suit) and three are not. That gives us one instance of $\frac{4!}{(1!)(3!)} = 4$ for each card in the straight.

Using the product rule, we get $10 \cdot 4^5 = 10240$.

10.5 Inclusion-Exclusion Principle

The inclusion-exclusion principle concerns the calculation of the number of elements in the union of several finite sets. The simple case of two sets was already touched upon in Section 6.7. As can be seen from the Venn diagram in Figure 31, the number of elements in $A \cup B$ is given by $n(A \cup B) = n(A) + n(B) - n(A \cap B)$, where $n(X)$ is a function that returns the number of elements in set X. The intersection needs to be subtracted since it is added in twice (once as part of A and again as part of B).

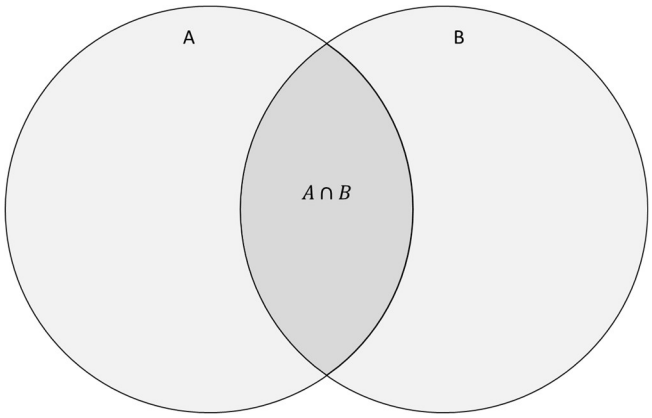


Figure 31. Inclusion-Exclusion for Two Sets

The situation for three sets (A, B and C) is a bit more complicated. Consider the Venn diagram in Figure 32. If we just add the number of elements in A, B and C together, the pairwise intersections are counted two times each and the $A \cap B \cap C$ is counted three times. To fix this, subtract each of the pairwise intersections. However, this removes the number of elements in $A \cap B \cap C$ three times. So, we need to add back the number of elements in $A \cap B \cap C$. The final result is

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

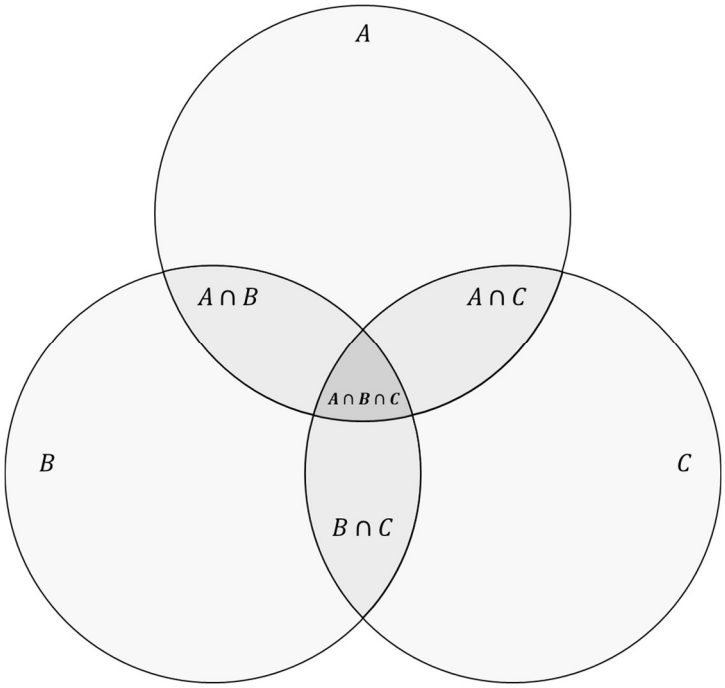


Figure 32. Inclusion-Exclusion for Three Sets

The formula for the number of elements in the union of r sets follow the same alternating pattern of additions and subtractions. It is basically a complex bookkeeping problem.

Theorem 10-2 (Inclusion-Exclusion Principle) $n(A_1 \cup A_2 \cup \ldots \cup A_r) = s_1 - s_2 + \cdots + (-1)^{r-1}s_r$, *where* $s_k = \sum n\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right)$ *for* $k = 1,2, \ldots, r$.

If Ω *represents the universe of discourse, then an alternate form of the principle can be written as:* $n(-A_1 \cap -A_2 \cap \ldots \cap -A_r) = n(\Omega) - s_1 + s_2 + \cdots + (-1)^r s_r$ *where* $-A_i$ *is shorthand notation for* $\Omega - A_i$.

The notation for s_k is a bit complex and requires additional explanation. In others words, we first add the number of elements in each A_i for $i = 1,2, \ldots, k$, i.e., $s_1 = n(A_1) + n(A_2) + \cdots + n(A_r)$. Next, we subtract the number of elements for the intersection of all possible pairs, i.e.,

$$s_2 = -[n(A_1 \cap A_2) + n(A_1 \cap A_3) + \cdots + n(A_1 \cap A_r) + n(A_2 \cap A_3) + n(A_2 \cap A_4) + \cdots + n(A_2 \cap A_r) + \cdots + n(A_{r-1} \cap A_r)].$$

After that, the number of elements for all possible intersections of triplets is added. The process alternates between including (adding) and exclusion (subtracting).

Proof: The approach taken here is to show that if $x \in \mathrm{A}_1 \cup \mathrm{A}_2 \cup \ldots \cup \mathrm{A}_\mathrm{r}$ appears in exactly k of the A_i (with k being any number between 1 and r) then the count of the number of appearances of x on left-side of the equation in the theorem statement is the same as the count of the number of appearances of x on the right-side. This would imply the equation is true.

Clearly x is counted once in the union of the A_i regardless of the value of k. So, the count on the left-side of the equation is 1. We are left to show that the count on the right-side of the equation for the number of appearances of x is also 1.

First, note that all intersections of more than k sets could not contain x since we stipulated that x is in exactly k of the A_i and not in the others (so zero appearances of x in this case). Further, all intersections in which one of the sets does not contain x also implies a zero count. The other cases are as follows:

- Of the k sets that do contain x, there are $k = \binom{k}{1}$ ways to select one at a time and so, x is counted $\binom{k}{1}$ times in this case.
- Of the k sets that do contain x, there are $\binom{k}{2}$ possible pairwise intersections and so, x is counted $\binom{k}{2}$ times in this case.
- Of the k sets that do contain x, there are $\binom{k}{3}$ possible ways to select 3 sets for intersection and so, x is counted $\binom{k}{3}$ times in this case.
- and so on.

Noting that the terms alternate between positive and negative in the formula, x is counted a total of $\binom{k}{1} - \binom{k}{2} + \binom{k}{3} \pm \cdots (-1)^{k-1}\binom{k}{k}$ on the right-side of the equation.

Using the binomial formula from Theorem 9-4, we have that

$$0 = (1-1)^k = \binom{k}{0} - \binom{k}{1} + \binom{k}{2} - \binom{k}{3} \pm \cdots (-1)^k\binom{k}{k} = 1 - \left[\binom{k}{1} - \binom{k}{2} + \binom{k}{3} \pm \cdots (-1)^{k-1}\binom{k}{k}\right].$$

Thus, $\binom{k}{1} - \binom{k}{2} + \binom{k}{3} \pm \cdots (-1)^{k-1}\binom{k}{k} = 1$ and we are done■

10.5.1 Divisibility

How many natural numbers less than or equal to 150 are **not** divisible by 2, 3 or 7?

This problem can be solved using the inclusion-exclusion principle. Define the following:

- $\Omega = \{1,2,3,\ldots,150\}$
- A is the set of natural numbers in Ω that are divisible by 2
- B is the set of natural numbers in Ω that are divisible by 3
- C is the set of natural numbers in Ω that are divisible by 7.

It is easy to see that $n(A) = 75, n(B) = 50, \text{and } n(C) = floor\left(\frac{150}{7}\right) = 21$, where the floor function returns the greatest whole number for a given argument.

What about $n(A \cap B)$, i.e., how many numbers less than or equal to 150 are divisible by both 2 and 3? In general, for prime numbers p and q, $p|a$ and $q|a$ if only if $pq|a$ (see Exercise #4 at the end of this section). So, 2 and 3 divide a number if and only if 6 does. Thus, $n(A \cap B) = floor\left(\frac{150}{6}\right) = 25$. Similarly, $n(A \cap C) = floor\left(\frac{150}{14}\right) = 10$ and $n(B \cap C) = floor\left(\frac{150}{21}\right) = 7$.

To determine $n(A \cap B \cap C)$, we note that for prime numbers p, q and r, $p|a, q|a$ and $r|a$ if and only if $pqr|a$. So, $n(A \cap B \cap C) = floor\left(\frac{150}{42}\right) = 3$.

From the alternate statement of the inclusion-exclusion principle, we have

$$n(-A \cap -B \cap -C)$$
$$= n(\Omega) - n(A) - n(B) - n(C) + n(A \cap B) + n(A \cap C) + n(B \cap C)$$
$$- n(A \cap B \cap C) = 150 - (75 + 50 + 21) + (25 + 10 + 7) - 3 = 43$$

Thus, there are 43 numbers less than or equal to 150 that are not divisible 2, 3 or 7.

10.5.2 Cryptography via Letter Substitutions

One approach for the secret encoding of messages is to define a table of substitutions for the letters in the alphabet. For example, Figure 33 depicts the key for a particular alphabet substitution cipher. The phrase "ATTACK AT DAWN" gets mapped to "KHHKMD KH NKQV". This type of code can be broken if a sufficient amount of encoded text is intercepted. In one approach to decipher the code, the codebreaker could look for letter frequencies in the coded text and match this to letter frequencies in the given language.

A	B	C	D	E	F	**G**	H	I	J	K	L	M	N	O	P	Q	R	**S**	T	U	V	W	X	Y	Z
K	L	M	N	O	P	**G**	A	B	C	D	E	F	V	W	X	Y	Z	**S**	H	I	J	Q	R	T	U

Figure 33. Alphabet Substitution Cipher

One desirable character of a substitution cipher is to not have any letter mapped to itself. In the example above, the two letters in bold (G and S) are fixed, i.e., left unchanged by the substitution. How many substitution ciphers are there that leave no letter fixed? This question can be answered using the inclusion-exclusion principle. We start by making the following assignments:

- Ω (universe of interest) – set of all possible rearrangements of the 26-letter English alphabet
- F_A – set of all rearrangements that leave the letter A fixed
- F_B – set of all rearrangements that leave the letter B fixed
- F_C – set of all rearrangements that leave the letter C fixed
- and so on.

First, note that $n(\Omega) = 26!$ (total number of ways of arranging all 26 letters of the alphabet).

For a given subset of j letters from the alphabet, there are $(26 - j)!$ arrangement of the alphabet that leave those j letters fixed. Using the labeling principle, there are $\frac{26!}{(26-j)!j!}$ ways of selecting j letters from the alphabet (labeling j of the letters as "fixed" and $26 - j$ of the letters as "not fixed"), which happens to be the same as the binomial coefficient $\binom{26}{j}$. So, there are $\binom{26}{j}(26 - j)!$ ways of rearranging the alphabet so that j of the letters are fixed, i.e., the number of rearrangements with j of the F_x properties at a time. Thus, we have the required input for the alternate form of the inclusion-exclusion formula, i.e.,

$$n(-F_A \cap -F_B \cap \ldots \cap -F_Z) = n(\Omega) - s_1 + s_2 + \cdots + (-1)^{26} s_{26}$$
$$= \sum_{j=0}^{26} \binom{26}{j} (-1)^j (26 - j)! = 148,362,637,348,470,135,821,287,825.$$

The first term in the above summation is $n(\Omega)$ (to see this, just plug-in $j = 0$).

10.5.3 Hat-check Problem

The hat-check problem entails the computation of how many ways the hats of n people can be removed and then redistributed to the n people with no person getting his or her hat. This is also known as the **problem of derangements**. The problem is identical to the letter substitution problem from the previous section when $n = 26$. For the case of n hats, we just replace 26 by n in the letter substitution problem to get the answer, i.e.,

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

For $n = 10$ people, the number of rearrangements (without anyone getting back their hat) is 1,334,961.

10.5.4 Surjective Mapping from One Finite Set to Another

Yet another variant of the hat-check problem is that of determining the number of surjective (onto) functions between a finite set A with m elements (numbered $1, 2, \ldots, m$) to another finite set B with n elements (numbered $1,2, \ldots, n$). Recall that surjective means that every element in the codomain (B in this case) has an element mapped to it from the domain (A in this case). Also, recall that a function maps each element in the domain to exactly one element in the codomain. However, it is allowed for a function to map several elements of the domain to the same element in the codomain. For the problem at hand, it must be that $m \geq n$; otherwise, $m < n$ forces at least one element of the domain to be mapped to more than one in the codomain but then we would not have a function in that case.

In order to use the inclusion-exclusion principle, define the following:

- Ω (universe of interest) – set of all possible functions from set A to B
- F_1 – set of all functions that do not map anything to 1 in the codomain
- F_2 – set of all functions that do not map anything to 2 in the codomain
- …
- F_n – set of all functions that do not map anything to n in the codomain.

For a function from A to B, each of the m elements of A can be mapped to any one of the n elements in B. Thus $n(\Omega) = n^m$. If we limit ourselves to subsets of size j from B, then the number of functions that do not map anything from A to those j elements in B is $(n-j)^m$. Using the labeling principle, there are $\binom{n}{j}$ ways of selecting a subset of size j from B. So, there are $\binom{n}{j}(n-j)^m$ functions from A to B such that j of the elements of B are missed (i.e., nothing is mapped to them from A by the function). This is the number of functions having j of the F_x properties at a time.

The number of surjective functions from A to B is equal to the number of functions that do not exhibit any of the properties $F_1, F_2, \ldots, F_n$, i.e.,

$$n(-F_1 \cap -F_2 \cap \ldots \cap -F_n) = n(\Omega) - s_1 + s_2 + \cdots + (-1)^n s_n$$
$$= \sum_{j=0}^{n} \binom{n}{j} (-1)^j (n-j)^m$$

Note that the function $n(A)$ concerning the number of elements in a set A and the variable n in the above equation are two different things.

10.6 Exercises

1. For a particular car model, one can choose among 7 exterior colors, 3 interior treatments and 5 packages of options. How many possible combinations are there? ***Hint**: Use the product rule.*
2. Find the number of proper divisors of 1,397,550, where proper divisor of a positive integer n is any divisor other than 1 and n. ***Hint**: The prime factorization of 1,397,550 is $2^3 5^2 7^1 11^3$, and remember to subtract 2, since 1 and n are not proper divisors.*
3. Given an integer N whose prime factorization is $p_1^{k_1} \cdot p_2^{k_2} \ldots p_r^{k_r}$, determine the number of proper divisors.
4. Prove that if A is any subset with 8 elements from the set {0,1,2,…,12}, then there are 2 elements of A whose sum is 12. ***Hint**: This is a pigeonhole principle problem. Let the pigeonholes be $H_1 = \{0,12\}, H_2 = \{1,11\}, H_3 = \{2,10\}, H_4 = \{3,9\}, H_5 = \{4,8\}, H_6 = \{5,7\}$ and $H_7 = \{6\}$. Let the elements of A be the pigeons.*
5. If 17 points are chosen at random in the interior of an equilateral triangle each side of which is 4 units long, show that some pair of points are within 1 unit of each other. ***Hint**: Apply the basic pigeonhole principle to Figure 34.*

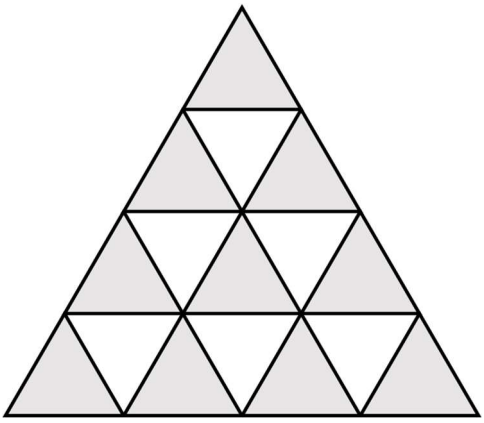


Figure 34. Subdivisions of an Equilateral Triangle

6. How many ways can the letters DDEFGGGHI be arranged? ***Hint**: Use the labeling principle.*
7. How many three-letter words can be formed from the letters DDGGG? ***Hint**: Just list all the possibilities.*
8. How many five-letter words can be formed from the letters AABBBDE? ***Hint**: This does not fit the pattern of any of the examples given in this section. There are four cases to be*

considered. The calculator at https://www.careerbless.com/calculators/word/index.php can be used to check your work. ***Answer****:250.*

9. From the set of integers from 1 to 30 inclusive, how many ways are there to select 3 even numbers and 4 odd numbers? ***Hint****: Use the labeling principle to determine the number of ways of separately selecting the even and odd numbers, and then use the product rule.*
10. How many 6-digit numbers have exactly one 2 and one 7? ***Answer****: Use the labeling principle to determine the number of ways of selecting one spot for the 2, one spot for the 7 and four spots for numbers other than 2 or 7, and then multiply times the number of ways of putting the other 8 numbers (i.e., 2 or 7) in the other 4 spots. This gives us* $\frac{6!}{1!1!4!}(8^4) = 122{,}880$.
11. Go to the Wikipedia article on "Poker probability" [42] and compute the frequency for a few of the hands listed, e.g., three of kind, or flush. ***Hint****: Use the product rule and labeling principle.*

11 Calculus

This section entails a very brief overview of calculus which will be needed to understand some of the topics covered later in this book, e.g., probability distribution functions.

[**Author's Remark**: I had hoped to avoid discussing calculus in this book, but found it impossible to cover continuous probability distributions without at least some use of calculus. Keep in mind that the typical calculus textbook is over 1000 pages and the summary that follows is but a few pages. Nevertheless, I hope to give the reader at least a high-level view of several fundamental concepts in calculus.

The American Institute of Mathematics provides links to many free mathematics textbooks, including several calculus books, see https://aimath.org/textbooks/approved-textbooks/.**]**

11.1 Limits

Simply put, a limit is the value that an expression (e.g., a function or sequence of numbers) approaches as the input to the expression approaches some value. In what follows, several basic examples of limits are given. For a good summary of the various properties of limits and a listing of limits for some common functions, see the Wikipedia article "List of limits" [39].

11.1.1 Example 1: f(x) = x/(x+10)

For example, the limit of $f(x) = \frac{x}{x+10}$ is 1 as x approaches infinity. In notation, this is written as $\lim_{x\to\infty} \frac{x}{x+10} = 1$. For large values of x, the numerator and denominator are almost the same (with ratio slightly less than 1). This formula can be proven but such details will not be covered in this overview. Continuing with this example, one can see that $f(x)$ approaches 1 (from above) as x approaches negative infinity. Thus, we can say $\lim_{x\to-\infty} \frac{x}{x+10} = 1$.

If one graphs $f(x)$ (see Figure 35), there are some additional points of interest. While $f(x)$ is not defined at $x = -10$ (division by zero is not allowed), it is still possible to define left and right-hand limits. Approaching $x = -10$ from the right, we can see that $f(x)$ takes on larger and larger negative values, as shown in Table 23. So, we have that $\lim_{x\to-10^+} \frac{x}{x+10} = -\infty$ where the notation $x \longrightarrow -10^+$ means that x approaches -10 from the right. If we approach -10 from the left, we get $\lim_{x\to-1^-} \frac{x}{x+10} = \infty$ where the notation $x \longrightarrow -10^-$ means that x approaches -10 from the left.

Table 23. Approaching x = -10 from the left

x	-9	-9.9	-9.99	-9.999	…
$f(x)$	-9	-99	-999	-9999	…

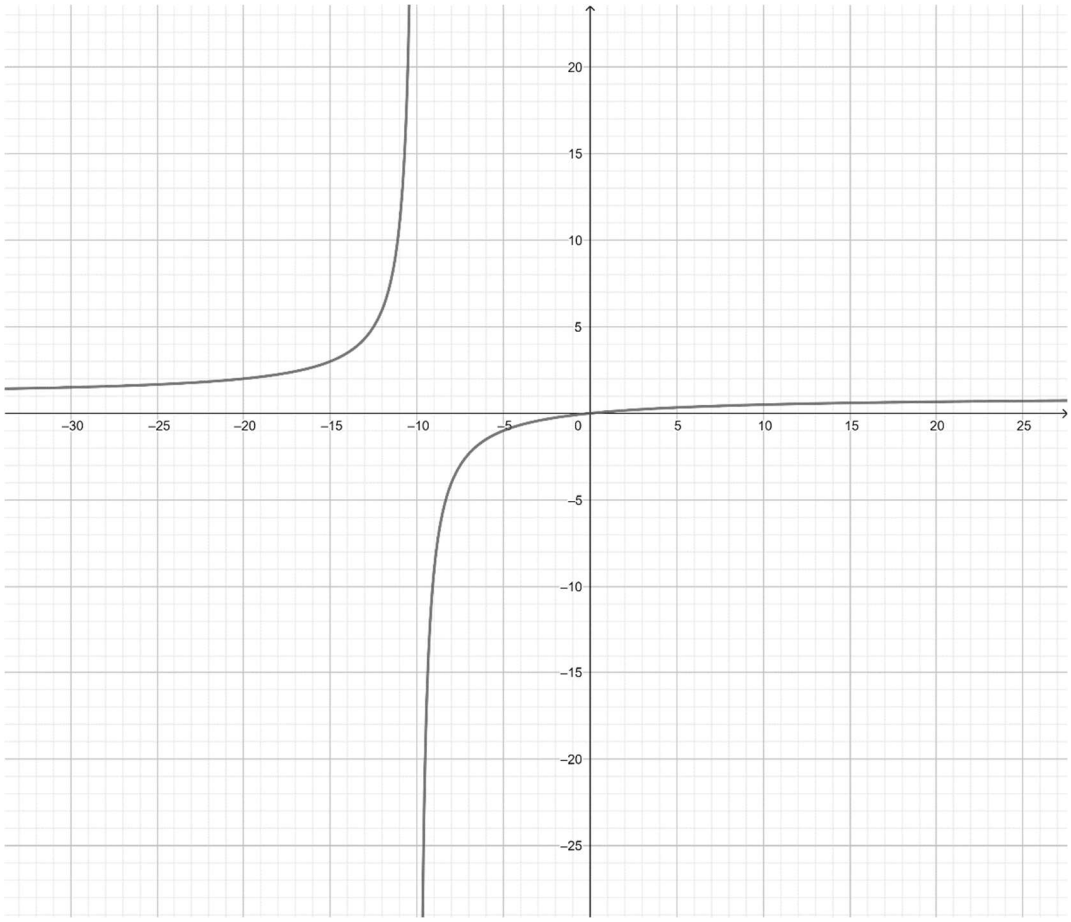


Figure 35. Limit Example concerning f(x) = x/(x+10)

11.1.2 Example 2: Missing Point in Straight Line

As another example, consider the function $g(x) = \frac{x^2-1}{x-1}$ which is undefined at $x = 1$ since division by 0 is undefined. When $x \neq 0$, we can simplify the function as follows: $g(x) = \frac{x^2-1}{x-1} = \frac{(x-1)(x+1)}{x-1} = x + 1$. The graph is basically that of $y = x + 1$ with a gap at the point $(1,2)$, shown as point A in Figure 36. Further, we have that $\lim_{x \to 1} \frac{x^2-1}{x-1} = 2$. So, $g(1)$ is undefined but the $\lim_{x \to 1} g(x) = 2$.

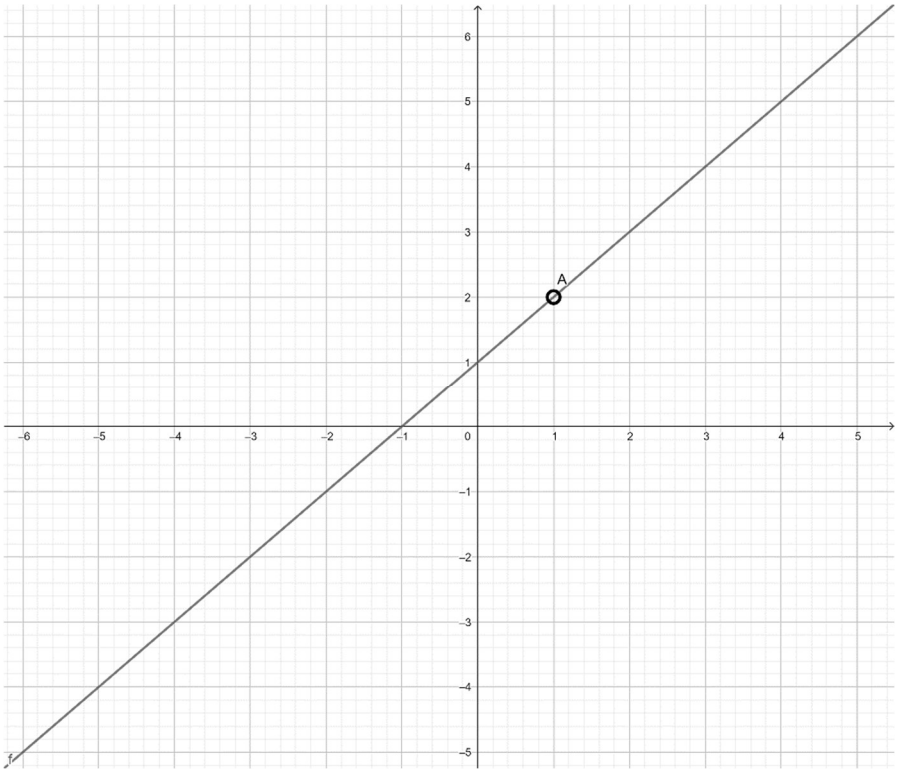


Figure 36. Limit Example for f(x) = x+1 but Undefined at x = 0

11.1.3 Example 3: Continuous Interest and Euler's Number

Consider an initial investment (P) in an interest bearing bank account. Let r be the nominal annual interest rate, n be the number of times per year interest is compounded, and t be the amount of time in years (or fraction of a year in increments of $\frac{1}{n}$) that interest is accrued. The principal grows as shown in Table 24.

Table 24. Compound Interest Formula

Time Period	Principal
0	P
1	$P + P\left(\frac{r}{n}\right) = P\left(1 + \frac{r}{n}\right)$
2	$P\left(1 + \frac{r}{n}\right) + \left(\frac{r}{n}\right)P\left(1 + \frac{r}{n}\right) = P\left(1 + \frac{r}{n}\right)^2$
3	$P\left(1 + \frac{r}{n}\right)^2 + \left(\frac{r}{n}\right)P\left(1 + \frac{r}{n}\right)^2 = P\left(1 + \frac{r}{n}\right)^3$
…	…
$\boldsymbol{nt}$	$P\left(1 + \frac{r}{n}\right)^{nt}$

What happens if interest is compounded an infinite number of times per year, i.e., $n \longrightarrow \infty$? It turns out that the formula actually converges, i.e.,

$$\lim_{n\to\infty} P\left(1+\frac{r}{n}\right)^{nt} = e^{rt}$$

The term e in the above formula appears frequently in mathematics and is known as **Euler's number** [40]. It is an irrational (actually transcendental) number whose value is approximately 2.7182818284. There are many ways to define e, with one of the most common definitions being $e = \lim_{n\to\infty} P\left(1+\frac{1}{n}\right)^{n}$. Another representation of e is given in the following theorem (which will be needed later in this book).

Theorem 11-1 $e = \sum_{i=0}^{\infty} \frac{1}{i!}$

Proof: From the binomial theorem, we have that

$$\left(1+\frac{1}{n}\right)^{n} = \sum_{i=0}^{n} \binom{n}{i}\left(\frac{1}{n}\right)^{i}$$

Next, note that

$$\binom{n}{i}\left(\frac{1}{n}\right)^{i} = \frac{n!}{(n-i)!\,i!}\cdot\frac{1}{n^i} = \left(\frac{1}{i!}\right)\frac{n(n-1)\ldots(n-i+1)}{n^i} = \left(\frac{1}{i!}\right)\cdot 1\cdot\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{i-1}{n}\right)$$

As $n \longrightarrow \infty$, the terms to the right of $\left(\frac{1}{i!}\right)$ in the above equation all converge to 1.

So, we have

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^{n} = \lim_{n\to\infty}\sum_{i=0}^{n}\binom{n}{i}\left(\frac{1}{n}\right)^{i} = \sum_{i=0}^{\infty}\lim_{n\to\infty}\binom{n}{i}\left(\frac{1}{n}\right)^{i} = \sum_{i=0}^{\infty}\frac{1}{i!}$$

which was to be proved■

The logarithm base e is known as the **natural logarithm** and is written as $ln\,x$ or (less commonly) as $log_e\,x$.

11.2 Differential calculus

Recall from basic algebra that a line of the form $y = mx + b$ has a slope of m. This means for every unit we change x, y changes by m. For example, take $y = 3x$. This is the equation for a straight line with slope 3. If we increase x from 1 to 2, then y changes from 3 to 6. The concept of a slope or rate of change is also important for more complex equations but differential calculus is required in such cases.

In short, the derivative of a function $f(x)$ is another function, denoted as $f'(x)$, that gives the slope of the tangent line to $f(x)$ at each point x in the domain of $f(x)$. An alternate notation for the derivative is $\frac{dy}{dx}$ where y is equal to some expression involving the variable x, e.g., $y = 3x^5 - 5x + \log_2 x$. While the details will not be covered in this book, it is worth noting that the formal definition of a derivative makes use of the concept of a limit.

Textbooks on calculus will typically devote many pages to the details of determining derivatives for various functions. However, for the task at hand in this book, the intent is to introduce the reader to the concept. Further, for many functions, it is possible to either lookup the derivative in a table (see, for example, the Wikipedia article on Differentiation rules [41]) or to use an online site to compute the derivative, e.g., Wolfram Alpha at https://www.wolframalpha.com or Symbolab at https://www.symbolab.com.

One of the most basic differentiation rules is captured in the following theorem:

Theorem 11-2 For functions of the form $f(x) = cx^a$ where c and a are constants, $f'(x) = cax^{a-1}$.

For example, the derivative of $f(x) = 3x^5$ is $f'(x) = 15x^4$. This means that at a given point $(x, 3x^5)$ on $f(x)$, the slope of the tangent line at that point is $15x^4$. An even simpler example is depicted in Figure 37. In the figure, the black curve is the graph of $f(x) = x^2$. From Theorem 11-2, we know that $f'(x) = 2x$. At the point $(2,4)$ on $f(x) = x^2$, the slope of the tangent line is $f'(2) = 4$. The graph of the tangent line is shown as the dashed line labeled h in the figure. At the point $(-1,1)$, the slope of the tangent line is -2 (see the dashed line labeled g in the figure).

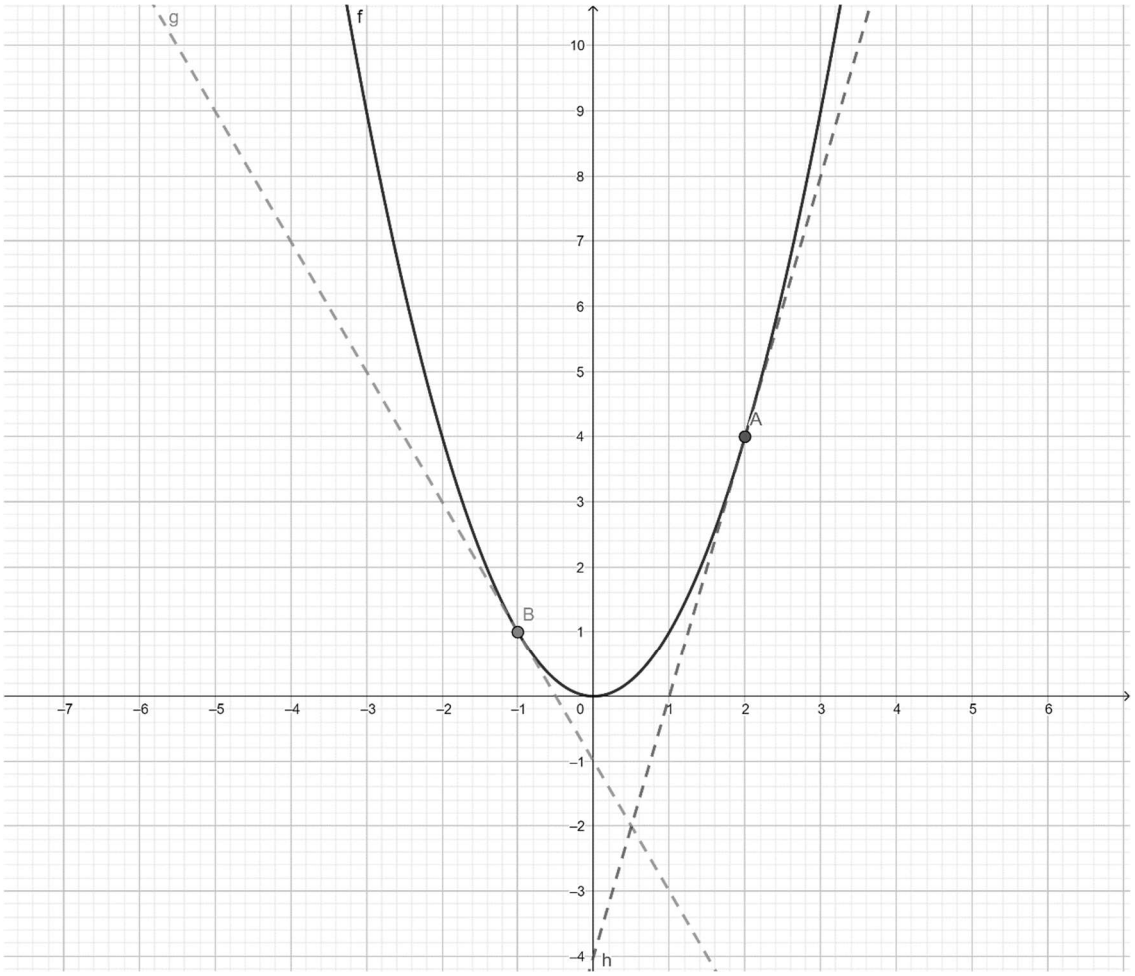


Figure 37. Tangents to a Parabola

There are several useful theorems that allow one to determine the derivative of a function which we state here without proof.

Theorem 11-3 The derivative of $f(x) = c^{ax}$ where a and c are constants, and $c > 0$ is $f'(x) = ac^{ax} \ln c$.

When $c = e$ (i.e., Euler's number) and $a = 1$ in the above theorem, we get that the derivative of e^x is itself, noting that $\ln e = 1$.

Theorem 11-4 The derivative of $f(x) = log_c x$ for $c > 0$ and $c \neq 1$ is $f'(x) = \frac{1}{x \ln c}$

When $c = e$, we get that the derivative of $\ln x$ is $\frac{1}{x}$

Theorem 11-5 (Derivative of a Linear Combination) If f and g are functions and $a, b \in \mathbb{R}$, then the derivative of $h(x) = af(x) + bg(x)$ is $h'(x) = af'(x) + bg'(x)$.

For example, the above theorem allows one to take the derivative of any polynomial given the previous result concerning the derivative of $f(x) = cx^a$. For example, the derivative of $f(x) = 3x^7 - 4x^5 + x^4 + 11x^{-2}$ is $f'(x) = 21x^6 - 20x^4 + 4x^3 - 22x^{-3}$.

Theorem 11-6 (Product Rule) If f and g are functions then the derivative of $h(x) = f(x)g(x)$ is $h'(x) = f(x)g'(x) + f'(x)g(x)$.

As an example of the product rule, consider the function $h(x) = x^3 \ln x$. Let $f(x) = x^3$and $g(x) = \ln x$ and use the product rule to get $h'(x) = x^3 \cdot \frac{1}{x} + 3x^2 \ln x = x^2(1 + 3 \ln x)$.

Theorem 11-7 (Quotient Rule)) If f and g are functions then the derivative of $h(x) = \frac{f(x)}{g(x)}$ is

$$h'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g^2(x)}$$

Theorem 11-8 (Chain Rule) The derivative of $h(x) = f(g(x))$ is $f'(g(x)) \cdot g'(x)$.

Recall that the composition of two functions was discussed in Section 8.2. For example, consider the function $h(x) = \ln(3x^7 - 4x^5)$. If we let $f(x) = \ln x$ and $g(x) = 3x^7 - 4x^5$, then $h(x) = f\big(g(x)\big)$. Further, this alternate representation of $h(x)$ allows us to find its derivative by using the chain rule, i.e.,

$$h'(x) = f'\big(g(x)\big) \cdot g'(x) = \frac{1}{g(x)} g'(x) = \frac{1}{3x^7 - 4x^5}(21x^6 - 20x^4) = \frac{21x^2 - 20}{3x^3 - 4x}, for\ x \neq 0$$

The one tricky part concerns $f'\big(g(x)\big)$ which is saying take the derivative of $f(x)$, which is $1/x$, and then replace x with $g(x)$.

11.3 Integral calculus

Consider the problem of determining the area of the shaded areas shown in Figure 38. In terms of notation this is written as follows: $\int_a^c f(x)\,dx$ and is read as "the integral of the function $f(x)$ from $x = a$ to $x = c$." The dx term just emphasizes that the integral is with respect to the variable x. This expression is referred to as a **definite integral** since an upper and lower bound is indicated. When computed, the integral gives the area between $f(x)$ and the x-axis, with the area above the x-axis being positive and the area below the x-axis being negative. So, for the function shown in Figure 38, $\int_a^c f(x)\,dx = A - B$. While the need for integration may seem obscure at this point, we will need to compute such areas regarding probability distributions in Section 12.9.

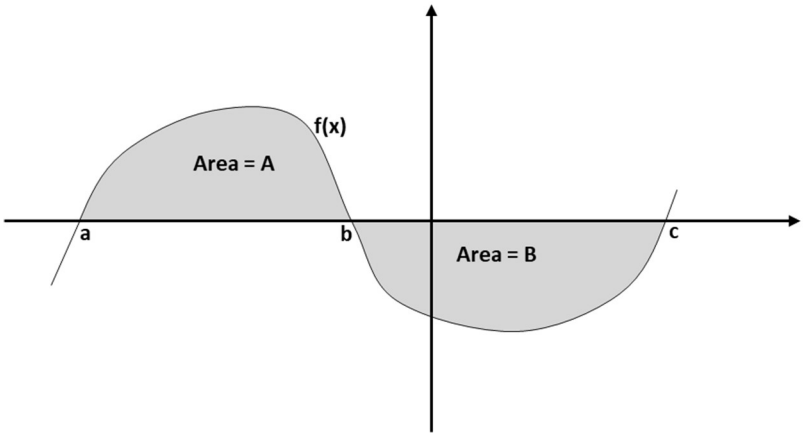


Figure 38. Area between f(x) and x-axis

It turns out that the computation of the integral is directly related to differential calculus, as noted in the following theorem.

Theorem 11-9 (Fundamental Theorem of Calculus) If $f(x)$ is a continuous function on the interval [a,b] and if $F(x)$ is such that $F'(x) = f(x)$ on the interval (a,b), then $\int_a^b f(x)\,dx = F(b) - F(a)$.

$F(x)$ is referred to as the **antiderivative** of $f(x)$ and is written as $\int f(x)\,dx$ which is referred to as an **indefinite integral** (since there are no upper or lower bounds).

The formulas stated in the previous section can be reversed to create statements about antiderivatives. For example, $\int x^n\,dx = \frac{1}{n+1}x^{n+1}$, for $n \neq 1$, and we can check this by just taking the derivative of the result. Further, there are entire books and online resources where one can look up the integrals for thousands of known functions. For example, WolframAlpha offers an online integral calculator at www.wolframalpha.com/calculators/integral-calculator.

11.3.1 Example: Integral of the Square Root of x

As an example, we determine the area under the function $f(x) = \sqrt{x} = x^{\frac{1}{2}}$ from $x = 4$ to $x = 16$ (see Figure 39). We first determine the antiderivative of $f(x)$, i.e.,

$$\int x^{\frac{1}{2}}\,dx = \frac{1}{(1+\frac{1}{2})}x^{\frac{1}{2}+1} = \frac{2}{3}x^{3/2} + c$$

Note that c is a constant and that the derivative of any constant is equal to 0. When the definite integral is computed, the constant cancels out.

Next, we use the Fundamental Theorem of Calculus to get

$$\int_4^{16} x^{\frac{1}{2}}\,dx = \left[\frac{2}{3}x^{3/2}\right]_{x=4}^{x=16} = \frac{2}{3}16^{\frac{3}{2}} - \frac{2}{3}4^{\frac{3}{2}} = \frac{2}{3}(4^3 - 2^3) = \frac{2}{3}(56) = 37.333\ldots$$

The notation $\left[\frac{2}{3}x^{3/2}\right]_{x=4}^{x=16}$ indicates the antiderivative and the two values of x that are to be substituted into the antiderivative in the following step.

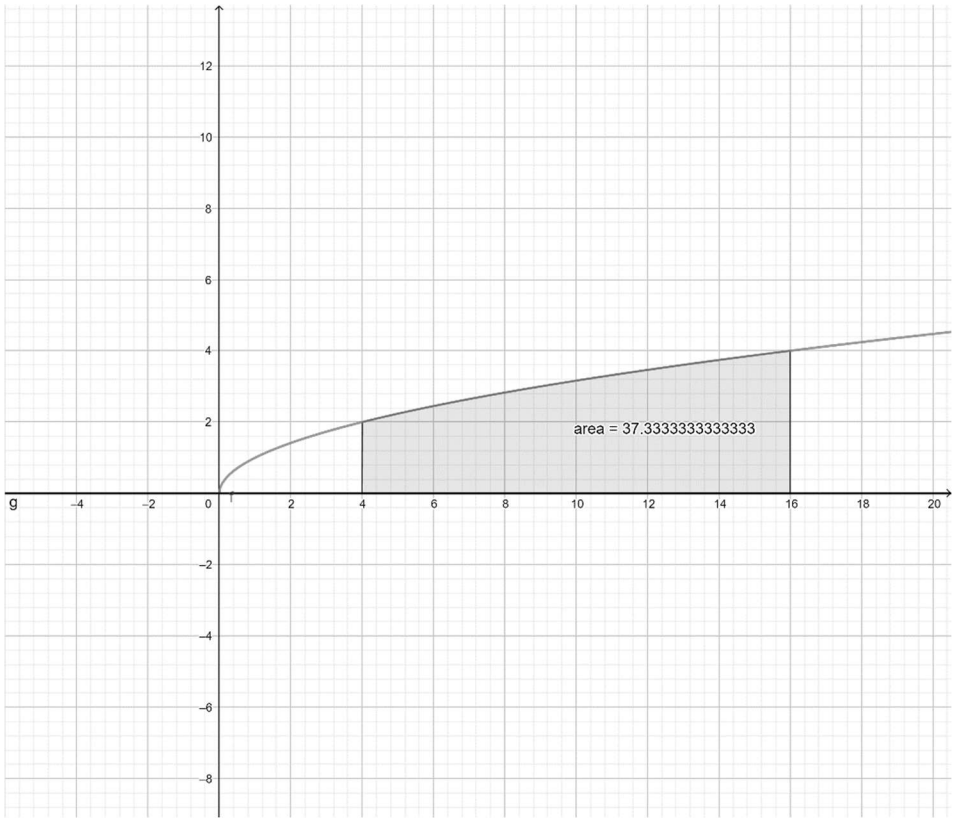


Figure 39. Area between a function and the x-axis

11.3.2 Example: Area Between Two Curves

In the Fundamental Theorem of Calculus, the area is actually being computed between two curves, i.e., $f(x)$ and the x-axis which is $g(x) = 0$. This can be generalized to any two continuous curves. For example, take the functions $f(x) = -\frac{x^2}{2} + 4$ and $g(x) = x^2 - 2$ (see Figure 40). To find the area between the two functions from $x = -2$ to $x = 2$, we first compute the antiderivative of $f(x) - g(x)$ as follows:

$$\int f(x) - g(x)\, dx = \int -\frac{3}{2}x^2 + 6\, dx = -\frac{x^3}{2} + 6x + c$$

We compute the definite integral from -2 to 2 as follows:

$$\int_{-2}^{2} f(x) - g(x)\, dx = \left[-\frac{x^3}{2} + 6x\right]_{x=-2}^{x=2} = -\frac{2^3}{2} + 6(2) + \frac{(-2)^3}{2} - 6(-2) == 16.$$

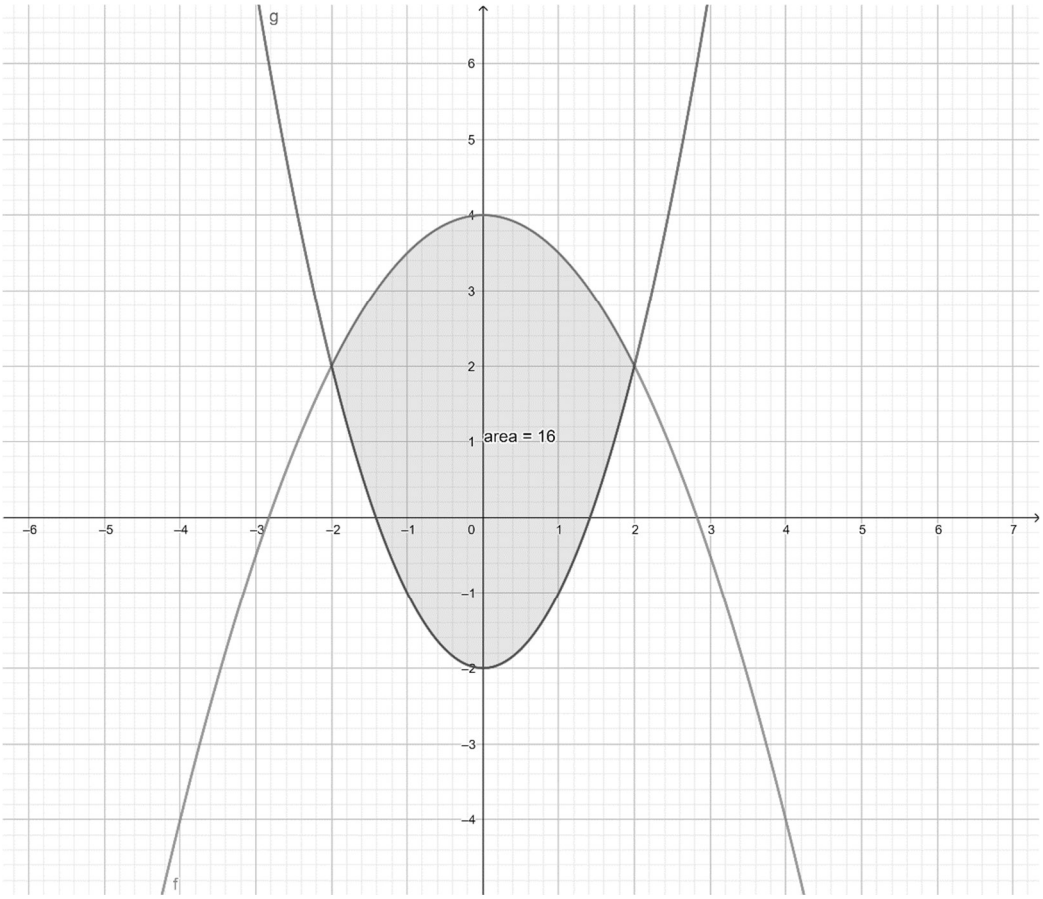


Figure 40. Area between Two Functions

11.4 Exercises

1. Find the derivative of $f(x) = x^3$ and the slope of the tangent line to $f(x)$ at the point $(2,8)$. Draw the graph of $f(x)$ and the tangent line at point (2,8). ***Hint**: Recall that a point (a, b) and a slope m uniquely determine a line. The associated formula is $y - b = m(x - a)$.*
2. Find the derivative of $f(x) = 7x^3 + 4x + \frac{1}{x^2} + \ln x$. ***Hint**: Use* Theorem 11-2*, Theorem 11-4 and Theorem 11-5*.
3. Find the tangent line to the circle $x^2 + y^2 = 1$ at the point $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. ***Hint**: Solve for y in terms of x and then take the derivative of y with respect to x.*
4. Use the product rule to find the derivative of $h(x) = (2x^5 + 3x)(5^x)$.
5. Use the quotient rule to find the derivative of $h(x) = e^x / \ln x$.
6. Use the chain rule to find the derivative of $f(x) = e^{x^2+5x}$.
7. Determine $\int_1^3 3x^4 - 2x^3\, dx$.
8. Determine $\int_1^\infty \frac{1}{x^2} dx$. ***Hint**: $\lim_{x\to\infty} \frac{1}{x} = 0$.*

9. Find z such that $\int_1^z x\,dx = 1$.
10. Find the anti-derivative of $\int x(x+3)^3\,dx$ using a table of known integrals. ***Hint**: Use Formula #7 at integral-table.com.*
11. Find the anti-derivative of $\int \frac{x^2}{(x-7)^3}dx$ using an online anti-derivative application (e.g., www.symbolab.com).

12 Probability

12.1 Definitions and Axioms

As a branch of mathematics, probability deals with determination of numerical likelihoods for various events. The likelihood values range from 0 to 1 where less likely events are assigned probabilities closer to 0 and more likely events have values closer to 1. The assignment of probabilities to events is entirely up to the modeler but for such assignments to be useful, the probabilities should reflect the likelihood of various events happening. For example, the probability of rolling a 2 with two dice is typically assigned the value $\frac{1}{36}$ since there are a total of 36 possible outcomes and only one outcome is 2 (i.e., both dice showing a 1).

The set (Ω) of all possible values of an experiment, e.g., rolling two dice, is called the **sample space** for the experiment. Each possible outcome of the given experiment is called a **sample point**. As with the assignments of probabilities, it is also up to the modeler to determine the sample space. For example, the experiment could be drawing 5 cards from a deck of playing cards. The sample space could be defined as the set of all possible 5-card selections which we know to be of cardinality $\binom{52}{5}$ from our work on combinatorics in the previous section. A particular 5-card draw is a sample point. Each sample point is assigned the same probability. Alternatively, the experiment might only reveal the sum of the values of the selected 5 cards (say the face cards are 10 points and the other cards are at face value) and not the individual cards. The sample points are now the possible sums of the 5-card draws (ranging from 6 to 50) and assigned probability will not be the same for each sample point since the likelihood of various sums are different.

Sample spaces can be classified as discrete (finite or countably infinite number of sample points), continuous (uncountably infinite number of sample points) or mixed (combination of discrete and continuous). Mixed sample spaces are not discussed further in this book. The experiment of rolling two dice has a discrete sample space. The experiment of randomly selecting a real number between 1 and 100 has a continuous sample space.

An **event** is a subset of the sample points from a sample space. Since events are sets, we can use the set terminology from Section 6.2 to talk about various combinations of sets, e.g., intersection and union. The probability of an event A is written as $P(A)$. In the context of this section, terms *event* and *set* are used interchangeably. Also, *experiment* and *sample space* are used interchangeably. The terminology is summarized schematically in Figure 41.

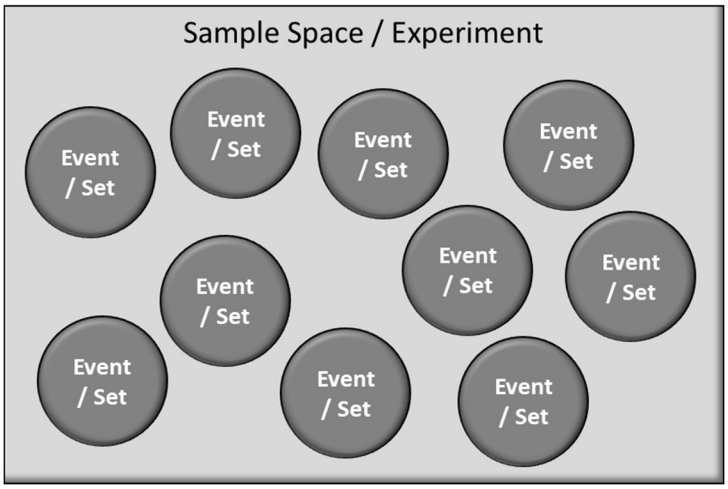


Figure 41. Summary of Probability Terminology

In the realm of probably, if A and B are events, and $A \cap B = \phi$ then A and B are said to be **mutually exclusive events** or equivalently **disjoint events**. For example, the event "it is raining now at location X" and the event "it is not raining now at location X" are mutually exclusive events.

The following axioms (assumptions) are posed. In the following, Ω is the entire sample space, and A and B are events within Ω (i.e., A and B are subsets of Ω).

- (Probability Axiom 1) For any event A, $P(A) \geq 0$.
- (Probability Axiom 2) $P(\Omega) = 1$ (If an experiment is carried out, then at least one of the sample points must occur.)
- (Probability Axiom 3) If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.
- (Probability Axiom 4) If $A_1, A_2, A_3, \ldots$ are mutually disjoint events, then

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Concerning the third axiom, this only works if A and B are disjoint. If A and B have sample points in their intersection, then some probabilities are counted multiple times. Recall the discussion concerning the inclusion-exclusion principle (that will come into play here for events that intersect).

From Axiom 3, we can prove a similar result for any finite number of mutually exclusive events (see Theorem 12-3). However, it should be emphasized that Axiom 4 cannot be proven from Axiom 3.

If we stipulate that $P(\phi) = 0$ (basically another axiom), then Axiom 3 can be derived from Axiom 4 if we set $A_3 = A_4 = \cdots = \phi$.

12.2 Approaches for Computing Probabilities

The probability of an event can be computed using either a combinatorial (a priori) approach or a statistical (a posteriori) approach. In the combinatorial approach, one determines the total number of possible outcomes of an experiment and the total number of ways a given event can occur. For example, there are $\binom{52}{5} = 2598960$ ways to draw five cards from a deck of playing cards and as we saw in Section 10.4, there are 3744 ways to draw a full house. Assuming each draw is equally likely

and using the combinatorial method, it would make sense to assign the probability of drawing a full house to be $\frac{3744}{2598960}$ which is about .001441. In the statistical approach, a given experiment is performed many times and the relative occurrence of various events is used to compute probabilities. For example, in the card example, one would simulate millions of 5-card draws to get estimates for how often a full house appears. The statistical approach may converge to the combinatorial approach depending on how closely the assumptions in the combinatorial model match reality. For example, if a coin is assumed to be fair, the probability of heads is .5. On the other hand, the coin could be biased and the statistical approach may reveal a probability of .55 for heads. Alternately, the coin itself may be unbiased but flipped in a way that favors tails and leads to an experimental probability of .6 for tails.

The basis for a statistical approach is discussed in Section 13.

12.3 Alternate Terminology for Likelihoods

The de facto standard for the expression of probabilities in mathematics is to use a range of values between 0 and 1. In other venues (e.g., gambling), other methods involving various types of odds are used. Some of the various types of odds are explained below. For a more complete description of odds, see the Wikipedia article on Odds [43].

12.3.1 Fractional Odds

For some types of gambling (e.g., horse racing), fractional odds are given. For example, a horse may be listed as 80:1 or 80/1 or 80-1 (read "eighty to one odds against winning"). This means that out of 81 chances the projection is that the horse has 80 chances of losing and 1 chance of winning, or in terms of probabilities, $\frac{1}{81}$ chance of winning and $\frac{80}{81}$ chance of losing. The odds are typically related to payouts for a bet. In the 80:1 example, someone betting $1 would expect to win $80 (plus return of their $1 bet) if their horse wins the race.

Odds can be stated for losing, winning or any other event (e.g., a horse coming in first, second or third). In general, $a{:}\,b$ odds of some event happening means that there is a projected probability of $\frac{a}{a+b}$ the event will happen and $\frac{b}{a+b}$ the event will not happen. For example, 2:5 odds that a football team will lose a particular game means the team is projected to win with probability $\frac{5}{7}$ and projected to lose with probability $\frac{2}{7}$. In this case, you would need to bet $5 on your team to win $2. If you wanted to bet on your team to lose, then a $2 bet would return $5.

Fractional odds are not always expressed using the lowest common denominator. For example, if there is a pattern of odds of 5:4, 7:4, 9:4 etcetera, then it would make more sense (in terms of readability) to express 3:2 as 6:4 (so as to fit the pattern of the other odds).

Saying that odds convert to some projected probability is not completely accurate. The odds are set by "the house" (some gambling related association such as a casino or horse racing establishment) to even out the betting among various options. Table 25 shows the odds from the 2019 Belmont Stakes horse race. The associated probabilities are shown in the right-most column (typically not shown by the gambling establishment). The probabilities may also be viewed as the amount of money that someone must bet to win $1. For example, if $\frac{1}{31}$ is bet on Joevia and Joevia wins, then the return is $30(\frac{1}{31})$ plus the initial amount of the bet $1/31$ for a total of $\frac{30}{31}+\frac{1}{31}=1$. So, in order to

be assured of a $1 return, it is necessary to bet on all horses (for the amount in the right-most column) for a total of $1.28. The difference is how the racetrack makes a profit.

Regarding the setting of odds, the racetrack (or other gambling establishment) will adjust the odds based on the amount of the bets on each option. For example, in the race noted in Table 25, if a lot of money started to be bet on the long shot (Joevia), the racetrack would quickly lower the odds on Joevia. The racetrack owners want to make money from the noted probability differential (which is guaranteed) and limit big losses from poorly set odds. In general, the more money that is bet on a given event, the more the gambling establishment will lower the odds and thus lower the payout per unit bet. This is regardless and completely independent of the gambling establishment's a priori view of the probability of a given event and in fact, they probably don't even care.

Table 25. Odds and Entries from 2019 Belmont Stakes

	Odds (not to finish 1st)	Probability of Winning
Joevia	30:1	$\frac{1}{31}$
Everfast	12:1	$\frac{1}{13}$
Master Fencer	8:1	$\frac{1}{9}$
Tax	15:1	$\frac{1}{16}$
Bourbon War	12:1	$\frac{1}{13}$
Spinoff	15:1	$\frac{1}{16}$
Sir Winston	12:1	$\frac{1}{13}$
Intrepid Heart	10:1	$\frac{1}{11}$
War of Will	2:1	$\frac{1}{3}$
Tacitus	9:5 or 1.8:1	$\frac{5}{14}$
Sum		1.28

Going in the other direction (i.e., given a probability and wanting convert to odds):

> If the probability for a given event is $p = \frac{r}{s}$, then the odds of winning are r:(s-r), noting that $s \geq r$ since $0 \leq p \leq 1$. For example, if the probability of an event happening is $\frac{1}{3}$ then the odds of that event happening are 1:2. If p is in decimal notation, this can always be converted to a fraction. If we are given $p = .77$, convert this to $\frac{77}{100}$ and then convert this to odds, i.e., 77:23.

12.3.2 Decimal Odds

Decimal odds are predominantly used (for betting) in continental Europe, Australia and Canada. The format is a numerical representation of the potential return of a bet, including the stake amount. For example, a $100 bet on Chelsea at 3.270 returns $3.270 \times 100 = \$327$, and the profit is $\$327 - \$100 = \$227$.

Decimal odds, as the name suggests, are simply $\frac{1}{p}$ where p is the projected probability of a given event. For example, consider the following decimal odds for two sports teams (to win):

Team A	4.0
Team B	1.25.

Team A has a projected probability of $\frac{1}{4} = .25$ of winning and Team B has a projected probability of $\frac{1}{1.25} = .8$ of winning. As before, we see the sum of the probabilities is greater than 1 (with the difference essentially being profits for "the house"). If the decimal odds for an event is x and an amount of y is bet, then a win results the amount of $xy - y$ in whatever currency is used for the bet. For example, if \$10 is bet on Team B and Team B wins, then the profit from the bet is $\$10(1.25) - \$10 = \$2.50$ whereas a \$10 bet on the underdog (Team A) results in a profit of \$30 (in the unlikely event Team A wins).

12.3.3 Moneyline Odds

Moneyline or American odds (used extensively for sports betting in the United States) are based on how much someone must bet to get a particular return. The basis is \$100. The following is an example Moneyline betting offer for an American football game:

Atlanta	-350
Jacksonville	+250

Usage of the minus sign in the manner employed here is unusual (at least with respect to mathematics). The -350 in the example means that a bet of \$350 is required to win \$100 (if Atlanta wins). The +250 in the example means that a bet of \$100 returns \$250 (if Jacksonville wins). Other bet amounts are possible and need to be adjusted proportionally.

In terms of fractional odds, -350 would be written as 100:350 (or 2:7) odds of losing, and +250 would be written as 250:100 (or 5:2) odds of losing.

In general, for a "minus odds" (say $-x$), the associated probability of winning is given by $\frac{x}{x+100}$ and for a "positive odds" (say y), the implied probability of winning is given by $\frac{100}{y+100}$. In the above example, the implied probability for Atlanta to win is $\frac{350}{350+1} = .778$ and the implied probability for Jacksonville is $\frac{100}{250+100} = .286$. Again, the sum of the probabilities is greater than 1.

Stating odds in terms chances of losing and stating probabilities in terms of chances of winning may seem unusual. However, this is fairly standard. See the various odds converters on the Internet, e.g., www.covers.com/editorial/HowToBet/OddsConverter or www.gamblingsites.org/sports-betting/odds-converter.

12.4 Basic Theorems

From the axioms stated in Section 12.1, several useful theorems can be proved. In what follows, reference to a set is understood to be a set of sample points within a sample space.

Theorem 12-1 For any set A, $P(A) \leq 1$.

Proof: A set and its complement comprise the entire universe of interest (Ω), i.e., $A \cup (\Omega - \mathrm{A}) = \Omega$. Further, A and $\Omega - A$ are disjoint. Using Axioms 2 and 3, we get

$$1 = P(\Omega) = P\big(A \cup (\Omega - \mathrm{A})\big) = P(A) + P(\Omega - \mathrm{A})$$

From Axiom 1, we know that $P(\Omega - \mathrm{A}) \geq 0$ and so $P(A) = 1 - P(\Omega - \mathrm{A}) \leq 1$■

Theorem 12-2 If $A \subset B$, then $P(A) \leq P(B)$ and $P(B - A) = P(B \cap \neg A) = P(B) - P(A)$.

Proof: Since $A \subset B$, we have that $B = A \cup (B - A)$. Noting that A and $B - A$ are mutually exclusive and using Axiom 3, we have that $P(B) = P(A) + P(B - A)$ which implies $P(A) \leq P(B)$ (noting that $P(B - A) \geq 0$ by Axiom 1) and $P(B - A) = P(B) - P(A)$■

Theorem 12-3 For disjoint sets $A_1, A_2, \ldots, A_n$, it follows that $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$.

Proof: The proof is by mathematical induction.

For $n = 2$, the result follows from Probability Axiom 3.

Assume the result holds of $n = k - 1$ and then show the result holds for $n = k$, i.e., $P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$. To that end, let $B = A_1 \cup A_3 \cup \ldots \cup A_{k-1}$ and note that A_k is disjoint from B. Using Axiom 3, we have

$$P((A_1 \cup A_2 \cup \ldots \cup A_{k-1}) \cup A_k) = P(B) + P(A_k)$$

and by the induction assumption we have that

$$P(B) = P(A_1 \cup A_2 \cup \ldots \cup A_{k-1}) = P(A_1) + P(A_2) + \cdots + P(A_{k-1})$$

Putting the prior two results together, we get $P(A_1 \cup A_2 \cup \ldots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$, which establishes the result for the case $n = k$■

Alternate Proof: In Probability Axiom 4, set $A_{n+1} = A_{n+2} = \cdots = \phi$ and we have the desired result■

For example, consider the sample space of all possible outcomes of rolling 2 dice. The sample space Ω is the set of all pairs (x, y) where x is the number rolled on Die #1 and y is the numbered rolled on Die #2. Let A_2 be the set of rolls that results in a two. There is only one sample point for A_2, i.e., $(1,1)$. Let A_3 be the set of rolls that results in a three, and so on, until A_{12}. The sets are disjoint, and thus, according to Theorem 12-3, we have that $P(\text{rolling a } 2, 3 \text{ or } 4) = P(A_2 \cup A_3 \cup A_4) = P(A_2) + P(A_3) + P(A_4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{1}{6}$.

As an example of where Theorem 12-3 does not apply, keep the same sample space from the previous example but define A to be the set of all rolls where Die #1 shows a 1, i.e., all sample points of the form $(1, x)$, and define B to be the set of all rolls that add to 7. Sets A and B are not disjoint since $(1,6)$ belongs to both sets, and Theorem 12-3 would give the incorrect result for

$P(A \cup B)$. However, it is still possible to determine a formula for $P(A \cup B)$ when A and B are not disjoint.

Theorem 12-4 For any two sets A and B (not necessarily disjoint), $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: We have the following equation by way of De Morgan's law:

$$A \cup (-A \cap B) = (A \cup -A) \cap (A \cup B) = \Omega \cap (A \cup B) = A \cup B.$$

However, A and $(-A \cap B)$ are disjoint, and so by Theorem 12-3, we have

$$P(A \cup B) = P(A \cup (-A \cap B)) = P(A) + P(-A \cap B) \qquad \text{(Equation 1)}$$

Next, we recast $-A \cap B$ as follows:

$-A \cap B$

$= (-A \cap B) \cup \phi$

$= (-A \cap B) \cup (B \cap -B)$ noting that $B \cap -B = \phi$

$= (B \cap -A) \cup (B \cap -B)$ commutative law

$= B \cap (-A \cup -B)$ distributive law in reverse

$= B \cap -(A \cap B)$ De Morgan's law

$= B - (A \cap B)$ by definition of set difference

Since $(A \cap B) \subset B$, we can use Theorem 12-2 to get

$$P(-A \cap B) = P(B - (A \cap B)) = P(B) - P(A \cap B) \quad \text{(Equation 2)}$$

Substituting Equation 2 into Equation 1 gives the desired result■

Theorem 12-4 can be applied to the previous dice example where Theorem 12-3 did not apply. We have that

$$P(\text{rolling a 1 on Die \#1 or rolling a total of 7 on both dice})$$
$$= P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{6}{36} - \frac{1}{36} = \frac{11}{36}.$$

Theorem 12-4 should be reminiscent of the inclusion-exclusion principle for the number of elements in the union of two (not necessarily disjoint) sets. This is not a coincidence. As noted in the Wikipedia article on the Inclusion-Exclusion Principle [44]:

> "As finite probabilities are computed as counts relative to the cardinality of a probability space, the formula for the principle of inclusion–exclusion remains valid when the cardinalities of the sets are replaced by finite probabilities. More generally, both versions of the principle can be put under the common umbrella of measure theory."

Probabilities and number counts are both measures of sets, where a measure μ is a function from a collection of sets to the real numbers that has three properties

- $\mu(A) \geq 0$ for every set A in the collection
- $\mu(\phi) = 0$

- $\mu(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ for all countable collections of pairwise disjoint sets in the collection.

Given the above properties, it is possible to prove the inclusion-exclusion principle and many other theorems. Both probability and number counts are measures. This book does not cover measure theory, beyond what has just been said. We mention it here to make the point that this type of abstraction (i.e., defining a theory that covers several more specific theories) is common (and important) in mathematics. For an introduction to measure theory, see the Wikipedia article on "Measure (mathematics)" [45].

The following is a statement of the inclusion-exclusion principle for probabilities. The formula in Theorem 12-5 has the same structure as the formula in Theorem 10-2 with the probability function replacing the "number of elements in a set" function.

Theorem 12-5 For the countable collection of disjoint sets $A_1, A_2, A_3, \ldots$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \pm \cdots (-1)^{n-1} \sum_{i<\cdots<n} P\left(\bigcap_{i=1}^{n} A_i\right)$$

12.5 Independent Events

Generally speaking, a collection of events is independent if the occurrence of any subset of the events has no bearing on the probability of the occurrence of the other events. Consider the example of drawing one card from each of two separate decks of playing cards. The draw from one deck has no effect on the draw from the other, and so, these two events are considered independent. However, if one sequentially draws two cards from a single deck, then the first draw definitely limits the second since one possibility has been removed. In the second scenario, the first and second draws from the single deck of cards are dependent events.

More formally, a collection of events $A_1, A_2, \ldots, A_n$ is said to be **mutually independent** (or just "independent" for short) if the intersection (i.e., joint occurrence) of any subset of them has probability equal to the product of probabilities of the individual events. More precisely, for any subset of the given events (say $\mathrm{A}_{i_1}, \mathrm{A}_{i_2}, \ldots, \mathrm{A}_{i_s}$), we have that

$$P(\mathrm{A}_{i_1} \cap \mathrm{A}_{i_2} \cap \ldots \cap \mathrm{A}_{i_s}) = P(\mathrm{A}_{i_1})P(\mathrm{A}_{i_2}) \ldots P(\mathrm{A}_{i_s}).$$

There is also a weaker condition known as **pairwise independence**, where only pairs of events (say A and B) from a collection of n events have the property $P(A \cap B) = P(A)P(B)$.

It is also worth noting that mutually exclusive events are necessarily dependent unless all of them are zero probability events.

12.5.1 Example: Pairwise but Not Mutually Independent

It is possible for a collection of events to be pairwise independent but not independent as a group. The following example appears in the book Introduction to Mathematical Statistics [46]. Let A and B be two independent flips of a fair coin. Represent heads by 1 and tails by 0. Define a third event C which takes the value 1 if the two flips together result in exactly one head; otherwise, C takes the value 0.

- We have $P(A=1)=P(B=1)=\ P(A=0)=P(B=0)=1/2$ since we assume a fair coin. Since events A and B are independent, we can multiple the probability of each event to get $P(A=0,B=0)=P(A=0,B=1)=P(A=1,B=0)=P(A=1,B=1)=\frac{1}{4}$.
- In order for $A=0$ and $C=0$, we must have the event $B=0$. So, $P(A=0,C=0)$ is the same as $P(A=0,B=0)=\frac{1}{4}$. Similarly,
  - $P(A=0,C=1)=P(A=0,B=1)=\frac{1}{4}$
  - $P(A=1,C=0)=P(A=1,B=1)=\frac{1}{4}$
  - $P(A=1,C=1)=P(A=1,B=0)=\frac{1}{4}$.
- The possible outcomes for (A,B,C) are $(0,0,0),(0,1,1),(1,0,1)$ and $(1,1,0)$. Each of the outcomes has probability ¼ since the outcome of the triplet only depends on the outcome of the first two events.

The above analysis shows that all the pairwise intersections follow the rule for independence, i.e., $P(X\cap Y)=P(X)P(Y)$. So, the collection of events (A, B and C) are pairwise independent. However, for all possible outcomes of A, B and C, we have $P(A\cap B\cap C)=\frac{1}{4}$ and this is not equal to $P(A)P(B)P(C)=(\frac{1}{2})^3=\frac{1}{8}$. Thus, A, B and C are not mutually independent.

12.5.2 Example: Independent Card Selections

As an example, consider the event of drawing one card each from three separate decks of playing cards. What is the probability of drawing an Ace of Hearts from each deck? Let A, B and C be the event of drawing an Ace of Hearts from the first, second and third decks, respectively. Since these events are independent, we have that $P(A\cap B\cap C)=P(A)P(B)P(C)=(\frac{1}{52})^3\cong 0.000007112$. On the other hand, if the three draws were taken from the same deck, the events would not be independent and of course, it would be impossible to draw the Ace of Hearts more than once.

12.5.3 Example: Independent Basketball Free-Throw

As another example, consider someone shooting free-throws in basketball. Assume a given person's probability of making a shot is p and therefore, $1-p$ probability of missing. What is the probability that this person will make k free-throws in n attempts, i.e., k successes and $n-k$ misses? Further, let's assume the shots are independent. In this case, the probability of a given arrangement of k successes and $n-k$ misses is $p^k(1-p)^{n-k}$ but we want to know the probability over all possible arrangements of k successes and $n-k$ misses. Recall this is a problem from the section on combinatorics. We have k events labeled as "success" and $n-k$ events labeled as "miss" and by the labeling principle, there are $\frac{n!}{(n-k)!k!}=\binom{n}{k}$ possibilities. Using Theorem 12-3 and adding the probabilities across $\binom{n}{k}$ events each with probability $p^k(1-p)^{n-k}$, we get that the desired probability is $\binom{n}{k}p^k(1-p)^{n-k}$. This is a well-known result in mathematics. The events in this example are called Bernoulli trials and the above formula is known as the binomial distribution. The binomial distribution will be discussed further in Section 12.9.3.2.

12.6 Conditional Probability

In some cases, we are asked to compute the probability of an event A given some prior information B that affects the calculation. Such probabilities are known as **conditional probabilities**. The probability of some event A given that event B has occurred is written as $P(A|B)$. For example, the probably that a random person has a runny nose (event A) may be 0.1 but the probability that the same person has a runny nose given that he or she has a virus infection (event B) may be much higher, e.g., $P(A|B) = 0.9$.

The following definition of conditional probably is posed as an axiom:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Notice that if A and B are independent, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

which is what one would expect. Further, the conditional probability is defined to be zero when the denominator is zero. If the conditional information (B) has no bearing on the event (A) in question then $P(A|B) = P(A)$.

For example, what is the probability that a card drawn from a deck of cards is a king given that the card drawn is a royal (i.e., king, queen or jack)? Let A be the event of drawing a king and B be the event of drawing a royal card. We are looking for $P(A|B)$. Noting that $P(B) = \frac{12}{52}$ and $P(A \cap B) = \frac{4}{52}$, and using the definition of conditional probability, we get $P(A|B) = \frac{4}{52} \div \frac{12}{52} = \frac{1}{3}$ which makes sense intuitively since there is $\frac{1}{3}$ chance of drawing a king when the selection is restricted to royal cards.

As noted in the following theorem, the concept of conditional probability can be extended to a collection of n events in a sample space.

Theorem 12-6 For n events $A_1, A_2, \ldots, A_n$ from a sample space, the probability of the intersection is given by the formula

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Proof: First, note that the order of the events in the equation really does not matter. We could have just as easily started with A_3 and some other arrangement after that.

Secondly, the equation is intuitively clear. The theorem is just saying "if you want to compute the probability of n events, start with the probability of any one event happening, multiply by the conditional probability of another event happening given the first event, and so on."

Now for the proof (which we do by the first principle of finite induction):

For $n = 2$, we already have the formula being true by the definition of conditional probability.

Assume the theorem is true for $n = k - 1$ and show that this implies the theorem is true for $n = k$. So, by the induction hypothesis, we have

$$P(A_1 \cap A_2 \cap \ldots \cap A_{k-1}) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_{k-1}|A_1 \cap A_2 \cap \ldots \cap A_{k-2}) \quad (1)$$

Since the intersection of events is itself an event, then we can use the definition of conditional probability to say

$$P(A_k|A_1 \cap A_2 \cap \ldots \cap A_{k-1}) = \frac{P(A_1 \cap A_2 \cap \ldots \cap A_k)}{P(A_1 \cap A_2 \cap \ldots \cap A_{k-1})} \quad (2)$$

Multiplying the left-side of equation (1) by the right-side of equation (2), and the right-side of equation (1) by the left-side of equation (2) gives

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_{k-1}|A_1 \cap A_2 \cap \ldots \cap A_{k-1})$$

which shows the result is true for $n = k$ and so, the proof is complete■

12.6.1 Example: School Subject Preference

Suppose that 75% of the children in a school like Mathematics, and 40% like both Mathematics and Science. What percentage of those who like Mathematics also like Science?

Let M be the event of liking Mathematics and S be the event of liking Science. Thus, we are being asked to find $P(S|M)$. We are given $P(M \cap S) = .4$ and $P(M) = .75$. By the definition of conditional probability, we have

$$P(S|M) = \frac{P(M \cap S)}{P(M)} = \frac{4}{10} \div \frac{75}{100} = \frac{40}{75} = 8/15 \text{ or about } 53.3\%.$$

A visual representation of the problem is depicted in Figure 42. The question can be recast as "if we restrict the sample space to Mathematics, what percentage of students also like Science?"

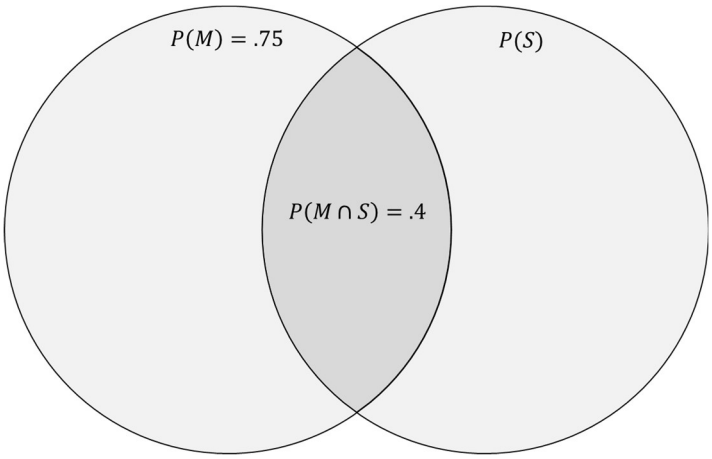


Figure 42. Venn Diagram for Student Preference Example

[Author's Remark: This is not technically a probability problem but rather, a percentage problem. However, the various axioms and theorems of probability also apply to percentages. It comes down to a matter of perspective. Percentages are typically the statement of facts. In the above example, we are told that 75% of the children like mathematics. The implication is that this number was measured exactly via a questionnaire or similar. If we said "there is a .75 probability that children like mathematics", the implication is that this is some sort of estimate (perhaps based on a sampling of a larger population).]

12.6.2 Example: Rolling an Octahedron Die

An octahedron die has 8 sides numbered from 1 to 8 (see Figure 43). The probability of rolling a given number is $\frac{1}{8}$, assuming the die is fair. The outcome of a roll is determined by the face that is touching the surface, i.e., the bottom of the die.

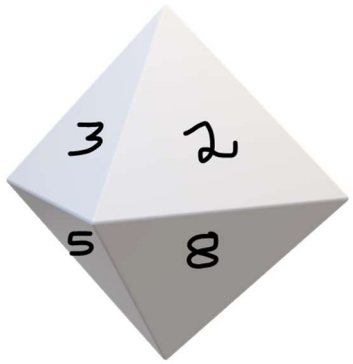


Figure 43. Octahedron Die

Let A be the event of rolling an odd number and B be the event of rolling at least 4. So, $A \cap B = \{5, 7\}$. Then we have $P(A) = \frac{4}{8}$, $P(B) = \frac{5}{8}$ and $P(A \cap B) = \frac{2}{8}$. Thus, $P(A|B) = \frac{2}{8} \div \frac{5}{8} = \frac{2}{5}$. This makes sense intuitively, since we are given that the roll is 4, 5, 6, 7 or 8, and two of the five numbers in that range are odd. We can also compute $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2}{8} \div \frac{4}{8} = \frac{1}{2}$ which also makes sense intuitively.

12.6.3 Example: Life Expectancy

Table 26 is an excerpt taken from the United States Life Tables [47] for the year 2017. The table shows the number of people still alive at a given age out of 100,000 live births. For the example that follows, we convert the number of people alive at a given age to a probability by inserting a decimal point before the number. For example, the number in box “Age=60 / Total” is converted to .88226, i.e., probability of being alive at age 60 is .882266. This is a statistical computation of probability.

Table 26. Number of survivors out of 100,000 born alive

Age	Total	Male	Female
0	100,000	100,000	100,000
1	99,422	99,370	99,477
5	99,326	99,261	99,393
10	99,268	99,199	99,341
15	99,191	99,107	99,280
20	98,937	98,749	99,134
25	98,466	98,071	98,883
30	97,872	97,235	98,543
35	97,163	96,284	98,083
40	96,321	95,196	97,493
45	95,275	93,903	96,697
50	93,797	92,105	95,543
55	91,538	89,365	93,768
60	**88,226**	85,344	91,162
65	**83,696**	79,838	87,596
70	77,697	72,785	82,637
75	69,418	63,524	75,344
80	57,839	51,095	64,591
85	42,382	35,439	49,264
90	24,560	18,687	30,222
95	9,361	6,070	12,383
100	1,894	971	2,697

Regarding the Total column in the above table, let T_N be the event of (at least) reaching the age of N. Focusing on the entries in bold red, we have that $P(T_{60}) = .88226$ and $P(T_{65}) = .83696$. Since $T_{60} \cap T_{65} = T_{65}$, we have that $P(T_{60} \cap T_{65}) = .83696$. Using the definition of conditional probability, we can compute probability of living to age 65 given that you are 60 which is

$$P(T_{65}|T_{60}) = \frac{P(T_{60} \cap T_{65})}{P(T_{60})} = \frac{.83696}{.88226} = .94866$$

As age increases, the probability of living another 5 years decreases. For example,

$$P(T_{90}|T_{85}) = \frac{P(T_{85} \cap T_{90})}{P(T_{85})} = \frac{.24560}{.42382} = .57949.$$

12.6.4 Pólya's Urn Model

This problem is named after famous mathematician George Pólya.

The Pólya Urn model entails an urn with r red balls and b black balls. The scheme is a multi-step process of drawing a ball, noting the color, and then returning the ball to the urn along with c more balls of the same color. The process is done a total of n times. The drawing of a ball of either color increases the probability of the same color being drawn in the next iteration. Pólya made an analogy with contagious diseases, where each case of a disease increases the probability of further cases.

Problem 1: Find the conditional probability that the second ball is red, given that the first ball is red.

> If the first ball is red, it is returned to the urn along with c additional red balls. So, there would be $r + c$ red balls and b black balls when the second ball is drawn. The probability of a red ball on the second draw would be $\frac{r+c}{r+c+b}$.

Problem 2: Find the probability that the first three drawings all result in red balls.

> Let A_n be the event that a red ball is drawn on the n^{th} draw. We are looking for $P(A_1 \cap A_2 \cap A_3)$ which we know to be $\mathrm{P}(\mathrm{A}_1)\mathrm{P}(\mathrm{A}_2|\mathrm{A}_1)\mathrm{P}(\mathrm{A}_3|\mathrm{A}_1 \cap \mathrm{A}_2)$ from Theorem 12-6. $P(A_1) = \frac{r}{r+b}$ from a simple combinatorial argument. In Problem 1, we have already computed $\mathrm{P}(\mathrm{A}_2|\mathrm{A}_1) = \frac{r+c}{r+c+b}$. If the first two draws are red, then we have $r + 2c$ red balls and b black balls. So, $\mathrm{P}(\mathrm{A}_3|\mathrm{A}_1 \cap \mathrm{A}_2) = \frac{r+2c}{r+2c+b}$ and so $\mathrm{P}(\mathrm{A}_1)\mathrm{P}(\mathrm{A}_2|\mathrm{A}_1)\mathrm{P}(\mathrm{A}_3|\mathrm{A}_1 \cap \mathrm{A}_2) = \frac{r(r+c)(r+2c)}{(r+b)(r+c+b)(r+2c+b)}$.

Problem 3: Find the conditional probability that the first ball is black, given that the second ball is black.

> We haven't yet developed the theory to solve this problem, but it is a good exercise to think about the solution. We'll come back to this problem in Section 12.8 concerning Bayes' theorem.

12.7 Law of Total Probability

Consider a container having 3 red and 7 blue marbles, and then remove one marble. What is the probability of selecting a blue marble on the second draw (assuming the first marble is not put back in the container)?

Let A be the event of interest, i.e., selecting a blue marble on the second draw. Let B_1 be the event of selecting a red marble on the first draw, and B_2 be the event of selecting a blue marble on the first draw. We have that

$$
\begin{aligned}
A &= A \cap \Omega \\
&= A \cap (B_1 \cup -B_1) \\
&= A \cap (B_1 \cup B_2) && \text{since } B_1 = -B_2\text{, i.e., } B_1 \text{ and } B_2 \text{ are mutually exclusive} \\
&= (A \cap B_1) \cup (A \cap B_2) && \text{by the distributive law}
\end{aligned}
$$

Thus, we have

$$P(A) = P(A \cap B_1) + P(A \cap B_2) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) = \frac{7}{9} \cdot \frac{3}{10} + \frac{6}{9} \cdot \frac{7}{10} = \frac{63}{90} = \frac{7}{10}$$

The concept illustrated in the previous example can be extended to a more general situation, as stated in the following theorem.

Theorem 12-7 (Law of Total Probability) For a collection of pairwise disjoint events $B_n, n = 1, 2, 3, \ldots$ whose union equals the entire sample space Ω, the probability of event A is given by

$$P(A) = \sum_{n=1}^{\infty} P(A \cap B_n) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n)$$

Proof: We have that

$A = A \cap \Omega$

$= A \cap (B_1 \cup B_2 \cup B_3 \cup \ldots)$ since the union of the B_i terms equal the entire sample space

$= (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \ldots \cup (A \cap B_n)$ distributive law.

Since the B_i terms are pairwise disjoint, we have that the $A \cap B_\mathrm{i}$ terms are pairwise disjoint (noting that $A \cap B_\mathrm{i} \subset B_\mathrm{i}$. Thus, we can use Theorem 12-3 to justify

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + \cdots + P(A \cap B_n)$$

and by the definition of conditional probability, we get

$$= P(A|B_1)P(B_2) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) + \cdots + P(A|B_n)P(B_n)$$

which as to be proved■

12.8 Bayes' Theorem

Given two events (A and B) from a sample space with the conditions that $P(A) \neq 0$ and $P(B) \neq 0$, it is straightforward to relate $P(A|B)$ to $P(B|A)$. From the definition of conditional probability, we have $P(A|B) = \frac{P(A \cap B)}{P(B)}$ and $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Solving the latter equation for $P(A \cap B)$ and substituting into the former equation, we get the following result:

Theorem 12-8 (Bayes' Theorem or Bayes' Rule) Given two events (A and B) from a sample space with the conditions that $P(A) \neq 0$ and $P(B) \neq 0$,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

It turns out that this is an extraordinarily useful equation. Entire books have been written on this and related topics, see [48], [49] and [50].

It is possible to extend Bayes' theorem by making use of the law of total probability.

Theorem 12-9 (Extended Bayes' Theorem) Let $A_k, k = 1,2, \dots, n$ be pairwise disjoint events whose union equals the entire sample space Ω, with $P(A_k) \neq 0, k = 1,2, \dots, n$. Let B be any event for which $P(B) > 0$, then for each $k = 1,2, \dots, n$, we have that

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_k)\,P(A_k)}$$

Proof: This follows directly from Bayes' theorem and the law of total probability, noting that $P(B) = \sum_{i=1}^{n} P(B|A_k)\,P(A_k)$ ■

12.8.1 Pólya's Urn Model

We now return to Problem 3 from Section 12.6.4, i.e., find the conditional probability that the first ball selected was black, given that the second ball selected is black. Let A be the event that the first ball selected is black and B be the event that the second ball selected is back. We are looking for $P(A|B)$. We have the following:

- $P(A) = \frac{b}{r+b}$
- $P(B|A) = \frac{b+c}{r+b+c}$ since if the first ball drawn from the urn is black, we add c black balls to the urn
- Noting that the event $-A$ means a red ball was selected on the first draw, we have that $P(-A) = \frac{r}{r+b}$ and $P(B|-A) = \frac{b}{r+b+c}$
- From the law of total probability,

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|-A)P(-A) \\ &= \frac{b(b+c)}{(r+b)(r+b+c)} + \frac{rb}{(r+b)(r+b+c)} \\ &= \frac{b}{(r+b)} \end{aligned}$$

Using Bayes' theorem, we get the desired result

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \left(\frac{b+c}{r+b+c}\right)\left(\frac{b}{r+b}\right)\left(\frac{r+b}{b}\right) = \frac{b+c}{r+b+c}$$

12.8.2 Colored Marbles in Three Containers

In this example, we have three containers with a mixture of red, white and black marbles in the amounts shown in Table 27.

Table 27. Bayes' Rule Example – Marbles in Containers

Container	Red	White	Black
1	4	2	1
2	1	3	3
3	5	2	5

A container is randomly chosen and one marble is drawn which turns out to be black. What is the probability that the marble came from Container 2?

Let A be the event that the selected marble is black. Let B_n be the event that Container n was chosen, where $n = 1,2 \text{ or } 3$.

Each container has an equal chance of being selected. So, $P(B_n) = \frac{1}{3}, \text{for } n = 1,2,3.$

From Table 27, we can determine that $P(A|B_1) = \frac{1}{7}, P(A|B_2) = \frac{3}{7}, \text{and } P(A|B_3) = \frac{5}{12}.$

Since $B_1 \cup B_2 \cup B_3 = \Omega$, we can use the extended Bayes' theorem as follows:

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{\sum_{i=1}^{3} P(A|B_n)\,P(B_n)} = \frac{\left(\frac{3}{7}\right)\left(\frac{1}{3}\right)}{\left(\frac{1}{7}\right)\left(\frac{1}{3}\right) + \left(\frac{3}{7}\right)\left(\frac{1}{3}\right) + \left(\frac{5}{12}\right)\left(\frac{1}{3}\right)} = \frac{36}{83}$$

[Author's Remark: The urn and ball examples are just a convenient abstraction. The general idea can be used in many situations. In the example above, the containers could represent departments in a company and the colors could represent different job functions. The mathematics is the same.]

12.8.3 Example: Sensitivity and Specificity

Bayes' theorem can be used to demonstrate the effect of false positives and false negatives in various tests, e.g., a test for a disease, or a determination of whether someone is guilty of a crime.

- **Sensitivity** is measured as the percentage of correctly identified positive determinations related to some test. So, a highly sensitive test would be expected to miss very few "positives", i.e., make a correct determination for situations that are actually positive for whatever is being tested. For example, in malaria tests, sensitivity is measured as the percentage of people who have a positive result on their malaria test and who, in fact, do have malaria.

- **Specificity** is measured as the percentage of correctly identified negative determinations related to some test. So, a highly specific test would be expected to miss very few "negatives", i.e., make a correct determination for situations that are actually negative for whatever is being tested. For example, in malaria tests, specificity is measured as the

percentage of people who have a negative result on their malaria test and who, in fact, do not have malaria.

If we let A be the event of actually having some characteristic and B be the event of testing positive for that characteristic, then

- Sensitivity is given by $P(B|A)$, and $P(-B|A)$ is the proportion of things that test negative but do have the characteristic, i.e., missed positives.
- Specificity is given by $P(-B|-A)$, and $P(B|-A)$ is the proportion of things that test positive but do not have the characteristic, i.e., missed negatives.

Consider a test for lead contamination (in residential dwellings) that is 95% sensitive and 85% specific. If .1 (or 1 in ten) houses have lead contamination, what is the probability that a randomly selected house with a positive test actually does have lead contamination?

Let A be the event that a randomly selected house does have lead contamination and let B be the event that a randomly selected house tests positive for lead contamination. We are looking for $P(A|B)$ and are given $P(A) = .1, P(B|A) = .95, P(-B|A) = .05, P(-B, -A) = .85$ and $P(B|-A) = .15$. If we can determine $P(B)$, then we can use Bayes' theorem to find $P(A|B)$. Using the law of total probability, we have that

$$P(B) = P(B|A)P(A) + P(B|-A)P(-A) = (.95)(.1) + (.15)(.9) = .23.$$

From Bayes' theorem, we have

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{(.95)(.1)}{.23} = .413043 \text{ or about } 41.3\%$$

Given the assumptions in this example, about 41.3% of the time a house with a positive lead contamination test will actually have lead contamination. So, this would not be a good test since $P(\neg A|B) = .587$ or about 58.7% of the houses that test positive for lead contamination actually do not have lead contamination. If used for determination of whether or not to purchase a house, a lot of purchases would be abandoned or unnecessary lead remediation would be done.

If it were possible to improve the specificity of the test, i.e., alter the value of $P(-B|-A)$, to what value would one need to increase the specificity for $P(A|B) > .9$?

To find a solution, we need to solve $\frac{(.95)(.1)}{(.95)*(.1)+(1-x)(.9)} > .9$ where $x = P(-B|-A)$. With some algebraic manipulation, we get that $P(-B|-A)$ needs to be greater than 98.83%.

12.9 Random Variables

12.9.1 Overview

In some instances, we are not interested in the sample points or events of an experiment per se, but rather on some numerical result based on sample points. As an example, consider a simple game of rolling two dice, where a player is awarded $2 on a $1 bet if the sum is odd, and $0 otherwise. In this example, there is a variable (call it X) which can take values $2 or $0. The value of X depends on chance and has associated probabilities. In general, a random variable is a numerical function defined on a sample space such that its specific value in any particular instance depends on chance and can be assigned a probability. In the dice game, $P(X = 0) = \frac{1}{2}$ and $P(X =$

$2) = \frac{1}{2}$. In general, a **random variable** is a function X that maps from a sample space S to a set of numbers Y. In function notation, we write $X: S \longrightarrow Y$.

A random variable that assumes a finite or countably infinite number of possible values is considered to be a **discrete random variable**, e.g., the draw of five cards from a deck of playing cards. A random variable that can take any value within a given interval is considered to be a **continuous random variable**, e.g., the temperature outside at a given location.

In what follows, Section 12.9.2 provides examples of random variables. The details concerning discrete and continuous random variables are discussed in separate subsections because the associated formulas for expected value (weighted average) and variance for the weighted average differ (see Sections 12.9.3 and 12.9.4, respectively). The discussion in each of these two subsections starts with some simple examples and is then followed by an overview of several commonly used probability distribution functions. An extensive list of discrete and continuous probability distributions functions can be found in the Wikipedia article entitled "List of probability distributions" [52].

12.9.2 Examples

12.9.2.1 Example 1: Rolling Two Dice

Let the sample space S be the outcomes of rolling two dice, i.e., $S = \{(1,1), (1,2), \ldots, (6,6)\}$. There are 36 elements in S. The typical random variable X defined on this sample space maps the sample point (a, b) to $a + b$. For example, $X(2,5) = 7$.

The probability that a discrete random variable takes on (i.e., "maps to") a given value is also of interest. For example, $P(X = 7) = \frac{6}{36}$ for the dice rolling random variable. It is critical to note that 6 of the sample points map to $X = 7$, i.e., $(1,6), (6,1), (5,2), (2,5), (3,4), (4,3)$.

The probability imposed on the various values of a discrete random variable induces a function, known as a **Probability Distribution Function** (PDF). For the given random variable X, Table 28 shows the associated probability distribution function, i.e., $f(x)$. The numbers 1 and 13 are included just to emphasize that probability function is zero for all values other than the integers 2, 3, …, 13. In general, the probability distribution function maps from the codomain of the random variable X to some subset of the interval [0,1].

Table 28. PDF for Roll of Two Dice using Sum of Pairs Random Variable

x	1	2	3	4	5	6	7	8	9	10	11	12	13
$f(x) = P(X = x)$	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0

It is also common to specify or compute what is call the **Cumulative Distribution Function** (CDF) for a random variable, which is defined as the function $F(x) = P(X \leq x)$. The CDF for the discrete random variable in the dice roll example is shown in Table 29.

Table 29. CDF for Rolling Two Dice using the Sum of Pairs Random Variable

x	1	2	3	4	5	6	7	8	9	10	11	12	13	…
$F(x) = P(X \leq x)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36} = 1$	1	1

While mapping the given sample space to the sum of the two dice is common, it is not the only possible mapping from the sample space to a set of numbers. For example, one could define a different random variable such as $Y(a, b) = max(a, b)$. This leads to a different PDF, i.e., $g(y)$, for the random variable Y, as shown in Table 30.

Table 30. PDF for Roll of Two Dice using Max

y	0	1	2	3	4	5	6	7
$g(y) = P(Y = y)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	0

The CDF for random variable Y is shown in Table 31.

Table 31. CDF for Roll of Two Dice using Max

y	0	1	2	3	4	5	6	7	…
$G(y) = P(Y \leq y)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{36}{36} = 1$	1	1

12.9.2.2 Example 2: Poker Hands

Let the sample space S be the possible outcomes from a draw of 5 cards from a deck of playing cards. Assume the various hands are ranked according to the following rules which define the random variable X:

- $X(royal\ flush) = 10$; Ten, Jack, Queen, King, and Ace, all of the same suit
- $X(straight\ flush) = 9$; Five consecutive cards of the same suit
- $X(four\ of\ a\ kind) = 8$
- $X(full\ house) = 7$; Combination of a three of a kind and a pair
- $X(flush) = 6$; Five cards of the same suit
- $X(straight) = 5$; Five consecutive cards
- $X(three\ of\ a\ kind) = 4$
- $X(two\ pairs) = 3$
- $X(pair) = 2$
- $X(high\ card) = 1$; assumes no pair or higher

In this case S has $\binom{52}{5} = 2598960$ sample points and the codomain of the random variable X is $\{1,2,3, \ldots, 10\}$. Clearly, many elements in the domain get mapped to the same element in the codomain. So, the random variable is not injective. However, it is surjective since for every element in the codomain there is at least one element from the domain S that gets mapped to it.

What is the probability that $X = 8$? This is a combinatorial problem where we need to divide the number of ways of getting four-of-a-kind by the total number of 5-card hands. To determine the number of ways of getting four-of-a-kind, we do the following:

- Use the labeling principle to determine the number of ways of selecting a rank. Label one of the ranks as "selected" and the other 12 ranks as "not selected". This gives a total $\frac{13!}{(12!)(1!)} = 13$ possibilities. This leaves 48 cards for selection as the 5th card in the hand.
- To determine the number of ways of selecting the 5th card in the hand, use the labeling principle. Label one card as "5th card in hand" and the other 47 cards as "not the 5th card in the hand." This gives $\frac{48!}{(47!)(1!)} = 48$ possibilities.
- Use the product rule on the above calculations to get $13 \cdot 48 = 624$ ways of getting four-of-a-kind.

So, $P(X = 8) = \frac{624}{2598960} = \frac{1}{4165}$ is the probability of four-of-a-kind in a 5-card poker hand.

12.9.2.3 Example 3: Continuous Random Variable concerning Temperature

As noted, a random variable is classified as continuous if it can assume all possible values in a given range of values. For example, assume the historical maximum daily temperature range for a particular city in the month of July is between 65 and 95 degrees Fahrenheit. It is reasonable to assume, for a given day in July, the maximum temperature in this city will almost definitely be between 60 and 100 degrees (allowing for possible new records). In this example, the random variable X is the maximum daily temperature and the sample space includes all values from 60 to 100. We can ask questions such as "what is the probability that the maximum temperature on a given day is between 80 and 85 degrees?" This can be written as $P(80 < X < 85)$. Unlike a discrete random variable, we don't have any point probabilities to sum. For example, in the dice example, to get the probability of a result between 6 and 8 (inclusive), we simply add the probabilities of $X = 6, X = 7$ and $X = 8$. In the continuous case, we use a curve to represent a probability distribution (rather than a set of point probabilities) where the area between the curve and the x-axis between two sample points is designed in such a way to give the probability that the random variable falls between the two sample points.

To illustrate the point, we assume all outcomes for the temperature in our example are equally likely in the range of 60 to 100. This is modeled with the probability density function $p(x) = 1/40$. The probability between two sample points in the given range is computed by determining the area bounded by the probability density function, the x-axis and the vertical lines going through the two points. So, $P(80 < X < 85)$ is the area between $p(x) = 1/40$ and the x-axis (black horizontal line in the figure) from 80 to 85, which is the area of the shaded rectangle shown in the Figure 44, i.e., 0.125. This is an example of what is called the uniform distribution (i.e., a constant value over an interval).

Further, notice that the area between $p(x)$ and the x-axis over the entire range (60 to 100) is exactly 1. This is necessarily true for every continuous probability density function. In Section 12.9.4, we will consider some typical probability density functions that are not simple horizontal lines. For these probability density functions, we need to compute definite integrals to determine the areas representing probabilities.

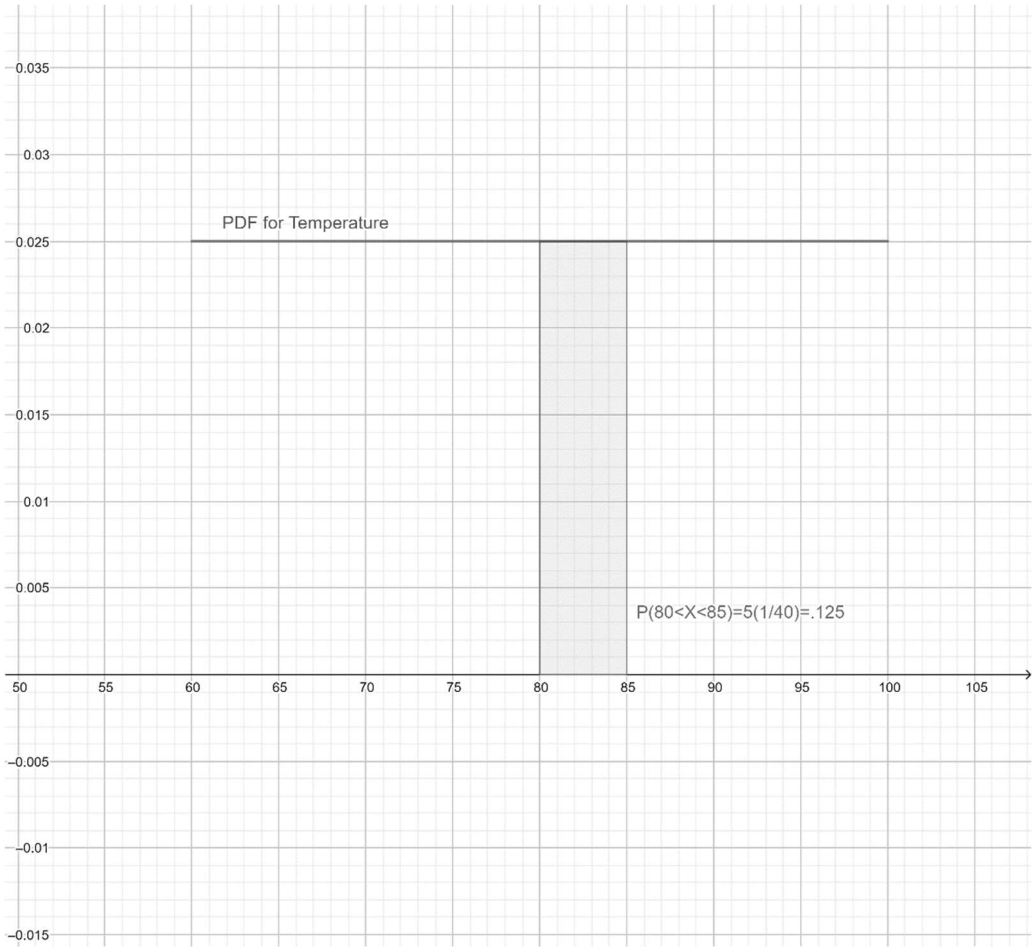


Figure 44. Example of Probability Computation for a Continuous Random Variable

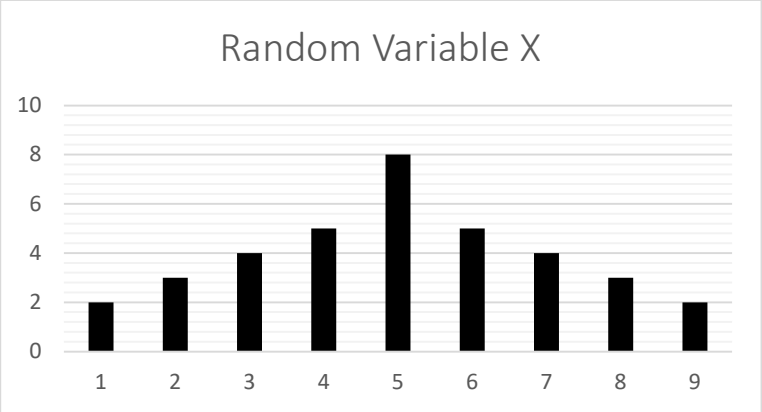
12.9.3 Discrete Random Variables

12.9.3.1 Definitions and Some Examples

In this section, various measures that summarize the distribution of discrete random variables are discussed. To facilitate the discussion, the following three examples are used. All three of the PDFs have the same weighted average (expected value) but their dispersions (spreads) are different.

Table 32. PDF for Random Variable X

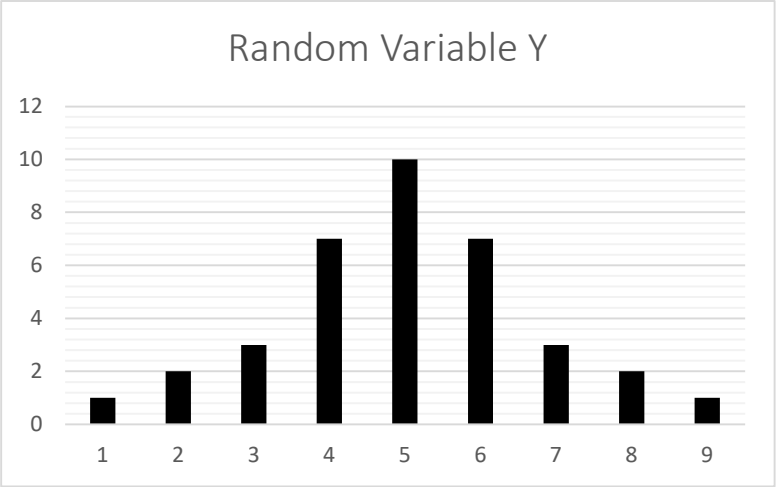
x	1	2	3	4	5	6	7	8	9
$f(x) = P(X = x)$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{8}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$



The PDF for random variable Y is more concentrated around the center value of $Y = 5$ than is random variable X.

Table 33. PDF for Random Variable Y

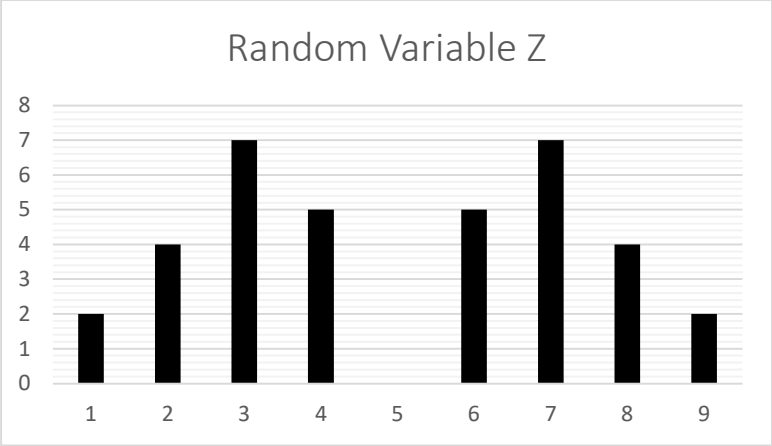
x	1	2	3	4	5	6	7	8	9
$f(x) = P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{7}{36}$	$\frac{10}{36}$	$\frac{7}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



The shape of the PDF for random variable Z is quite different from that of X and Y, but the weighted average is the same for all three (as is shown below).

Table 34. PDF for Random Variable Z

x	1	2	3	4	5	6	7	8	9
$f(x) = P(X = x)$	$\frac{2}{36}$	$\frac{4}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	0	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{4}{36}$	$\frac{2}{36}$



In order to analyze PDFs more accurately, several common measures are defined.

The **expected value** of a discrete PDF is basically the weighted average. If $f(x_i) = P(X = x_i)$ for $i = 1, 2, 3, \ldots$ then the expected value of X is defined as

$$E(X) = \sum_{i=1}^{\infty} x_i f(x_i) = x_1 f(x_1) + x_2 f(x_2) + x_3 f(x_3) + \cdots$$

The above formula also holds if the PDF has only a finite number of terms n (just replace ∞ with n in the summation). The Greek letter μ is often used as a shorthand for $E(X)$.

For the three example PDFs, we have that

$$E(X) = 1 \cdot \frac{2}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{4}{36} + 4 \cdot \frac{5}{36} + 5 \cdot \frac{8}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{4}{36} + 8 \cdot \frac{3}{36} + 9 \cdot \frac{2}{36} = \frac{180}{36} = 5$$

$$E(Y) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{2}{36} + 3 \cdot \frac{3}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{10}{36} + 6 \cdot \frac{7}{36} + 7 \cdot \frac{3}{36} + 8 \cdot \frac{2}{36} + 9 \cdot \frac{1}{36} = \frac{180}{36} = 5$$

$$E(Z) = 1 \cdot \frac{2}{36} + 2 \cdot \frac{4}{36} + 3 \cdot \frac{7}{36} + 4 \cdot \frac{5}{36} + 5 \cdot \frac{0}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{7}{36} + 8 \cdot \frac{4}{36} + 9 \cdot \frac{2}{36} = \frac{180}{36} = 5$$

So, as previously noted, all three of the PDF examples have the same expected value. However, their spread and shapes are different. In terms of spread, there is a standard measurement, i.e., standard deviation (from the expected value), which is based on something called variance.

The **variance** of a discrete PDF is a measure of how much the values of the random variable vary from the expected value. If $f(x_i) = P(X = x_i)$ for $i = 1, 2, 3, \ldots$ and $\mu = E(X)$, then the variance of X is defined as

$$Var(X) = \sum_{i=1}^{\infty} (x_i - \mu)^2 f(x_i) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + (x_3 - \mu)^2 f(x_3) + \cdots$$

The above formula also holds if the PDF has only a finite number of terms. While the proof is omitted here, it is true that $Var(X) = E(X^2) - \mu^2$ which is easier to use for computation (especially by hand). $Var(x)$ is usually abbreviated as σ^2.

Now, we can define the **standard deviation** of a PDF as $\sigma = \sqrt{Var(X)}$.

For the three examples,

$$E(X^2) = 1 \cdot \frac{2}{36} + 2^2 \cdot \frac{3}{36} + 3^2 \cdot \frac{4}{36} + 4^2 \cdot \frac{5}{36} + 5^2 \cdot \frac{8}{36} + 6^2 \cdot \frac{5}{36} + 7^2 \cdot \frac{4}{36} + 8^2 \cdot \frac{3}{36} + 9^2 \cdot \frac{2}{36} = \frac{265}{9}$$

and so, $Var(X) = E(X^2) - \mu^2 = \frac{265}{9} - 25 = \frac{40}{9}$ and $\sigma_X \cong 2.11$. (Note that the subscript X on the standard deviation is to distinguish it from the standard deviations of random variables Y and Z.)

$$E(Y^2) = 1 \cdot \frac{1}{36} + 2^2 \cdot \frac{2}{36} + 3^2 \cdot \frac{3}{36} + 4^2 \cdot \frac{7}{36} + 5^2 \cdot \frac{10}{36} + 6^2 \cdot \frac{7}{36} + 7^2 \cdot \frac{3}{36} + 8^2 \cdot \frac{2}{36} + 9^2 \cdot \frac{1}{36} = \frac{503}{18}$$

and so, $Var(Y) = E(Y^2) - \mu^2 = \frac{503}{18} - 25 = \frac{53}{18}$ and $\sigma_Y \cong 1.72$ which is smaller than σ_X. This is expected since the PDF for Y is less dispersed than the PDF for X.

$$E(Z^2) = 1 \cdot \frac{2}{36} + 2^2 \cdot \frac{4}{36} + 3^2 \cdot \frac{7}{36} + 4^2 \cdot \frac{5}{36} + 5^2 \cdot \frac{0}{36} + 6^2 \cdot \frac{5}{36} + 7^2 \cdot \frac{7}{36} + 8^2 \cdot \frac{4}{36} + 9^2 \cdot \frac{2}{36} = \frac{551}{18}$$

and so, $Var(Z) = E(Z^2) - \mu^2 = \frac{551}{18} - 25 = \frac{101}{18}$ and $\sigma_Z \cong 2.37$. This is even larger than the standard deviation for X but that only tells part of the story. More significant is the shape of the PDF for Z which is bimodal (2 peaks) unlike the unimodal distributions for X and Y. These three things, i.e., expected value, standard deviation and shape, are commonly used to describe PDFs.

12.9.3.2 Binomial Distribution

The binomial distribution was previously introduced as a solution to a problem in Section 12.5.3. It is the distribution for the number of successes in n Bernoulli trials where each trail has probability p of success and probability $1 - p$ of failure. For example, consider a baseball player that has a batting average of .333 (i.e., on average gets one hit in three times at bat). Each at bat (usually abbreviated as AB in baseball stats) is a Bernoulli trial with $p = .333$. The probability that the particular baseball player gets 40 or more hits in 100 at bats can be modeled with a binomial distribution.

In general, the PDF for a random variable X that is modeled with the binomial distribution is $f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ where n is the number of trials, k is the number of successes, and p is the probability of a success in a trial. From the binomial theorem (Theorem 9-4), we have that PDF sums up to 1, i.e.,

$$1 = [p + (1-p)]^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

Returning to the baseball example, the probability of 40 or more hits in 100 bats (when modeling this activity with a binomial distribution) is given by

$$P(X = 40) = \sum_{k=40}^{100} \binom{100}{k} (.333)^k (.667)^{100-k} \cong .06495$$

Computation of the expected value for the binomial distribution is as follows:

$\mu = E(X) = \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$ definition of expected value

$= \sum_{k=1}^{n} k\binom{n}{k} p^k (1-p)^{n-k}$ 1st term is 0, can start from $k = 1$

$= \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k}$ since $k\binom{n}{k} = n\binom{n-1}{k-1}$

$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$ since $n-k = (n-1)-(k-1)$

$= np \sum_{j=0}^{m} \binom{m}{j} p^j (1-p)^{m-j}$ letting $m = n-1, j = k-1$

$= np.$

Using a similar but longer and more complex computation than the above, one can determine that $E(X^2) = n^2p^2 + np(1-p)$. Details of the proof can be found at Proof Wiki in the article "Variance of Binomial Distribution" [55]. This implies that $Var(X) = np(1-p)$ and by definition, $\sigma = \sqrt{np(1-p)}$.

As an example, take $n = 10$ (i.e., 10 Bernoulli trials) with probability $p = .7$ of success on each trial. In this case, $\mu = 7$ and $\sigma \cong 1.4491$. The graph of the PDF is shown in Figure 45. The number at the top of each column is the probability for a given number of successes out of 10 trials, e.g., the probability of 6 successes is .2001.

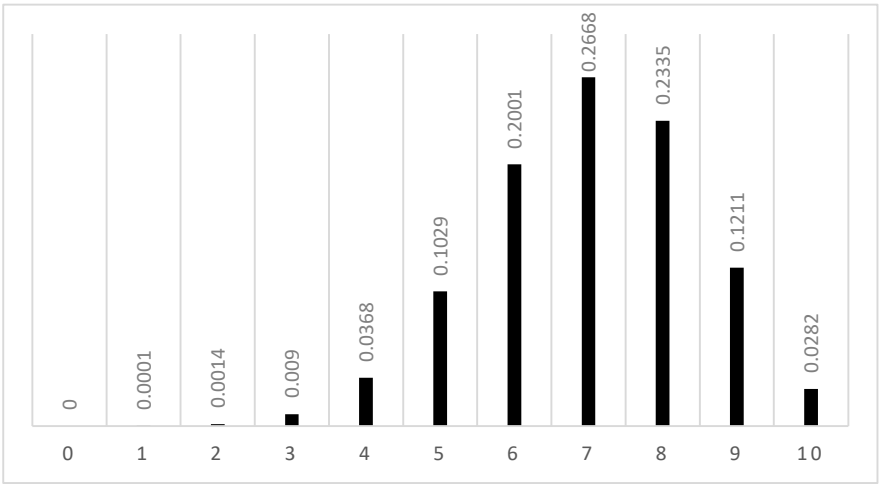


Figure 45. Graph of Binomial Distribution, n=10, p=.7

12.9.3.3 Geometric Distribution

The geometric distribution gives the probability for the number of Bernoulli trials needed to get an initial success. Using the baseball example from the previous section, the geometric distribution could be used to model the probability of a baseball player getting his first hit after some given starting point.

In general, the PDF for the geometric distribution is given by $f(k) = P(X = k) = (1-p)^{k-1}p$, where p is the probability of success for each Bernoulli trial and k is the number of the first successful trial. Using the result from Theorem 6-12, we see that the probabilities do sum to 1:

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=0}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1$$

The expected value of the geometric distribution is $\frac{1}{p}$ and the variance is $\frac{1-p}{p^2}$.

Table 35 shows probability values for the geometric distribution for $p = .1, .3 \text{ and } .5$ and for k from 1 to 7 (noting that k continues indefinitely with the probabilities tending to zero). The larger the value of p, the more quickly that the distribution approaches 0.

Table 35. Geometric Distribution Examples

k	1	2	3	4	5	6	7
p=.1	0.09	0.081	0.0729	0.06561	0.05905	0.05314	0.04783
p=.3	0.21	0.147	0.1029	0.07203	0.05042	0.03529	0.02471
p=.5	0.25	0.125	0.0625	0.03125	0.01563	0.00781	0.00391

The graph of the points from the geometric distributions in Table 35 are shown graphically in Figure 46. It is emphasized that the geometric distribution is discrete and the lines connecting the points in the figure are not part of the distribution but are just there to show the trend of the distributions.

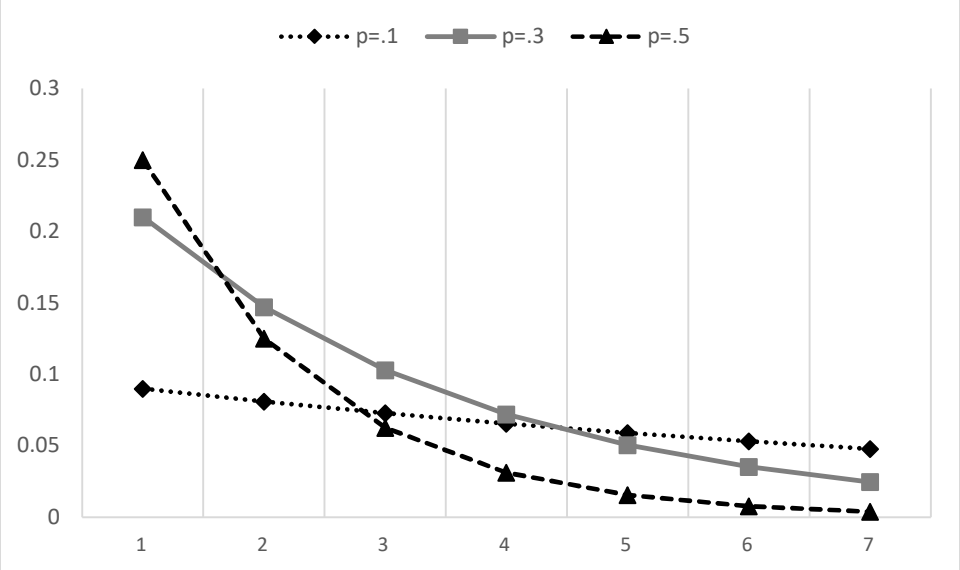


Figure 46. Graph of Several Geometric Distributions

12.9.3.4 Poisson Distribution

The Poisson distribution represents the probability of a given number of events (k) occurring in a **fixed** interval of time or space under the assumption that the events occur independently at a constant rate (λ). For example, the Poisson distribution is commonly used to model the arrival of entities into a queue (see, for example, the Wikipedia article on M/M/1 queues [59] where the arrival process has a Poisson distribution). Application of the Poisson distribution with regard to space (e.g., events occurring over some area or volume) is less common. Space-related examples include

- the occurrence of crater centers on the moon, see the article “On a Test of Randomness of Lunar Craters” [60]
- the number of misprints on a page of a book.

The PDF is very simple, i.e., $f(k) = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ where λ is the average number of events per fixed interval and k is the number of events. The value of k ranges from 0 to ∞.

It turns out that the Poisson distribution is the limiting case of the binomial distribution when the number of trials n approaches infinity, the probability of success p approaches 0, and np remains constant. For a proof of this fact, see the YouTube video entitled “Proof that the Binomial Distribution tends to the Poisson Distribution” [61]. This relationship leads to a useful interpretation of the Poisson distribution, which we exhibit in the following example. Consider a cash machine outside of a bank and the set of customers that typically use the machine. The probability that any particular customer will use the machine in say a given 15-minute period is very small. However, there are a large number of customers that do use the machine and so, the number of customers using the machine in a given 15-minute period is typically at a fixed rate (this can be verified via an observation experiment). Thus, usage of the cash machine in this example is a good candidate to be modeled with the Poisson distribution.

Getting back to the details of the Poisson distribution, the expected value is computed as follows:

$$\mu = E(X) = \sum_{i=0}^{\infty} \frac{i\lambda^i e^{-\lambda}}{i!} = \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{(i-1)!} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

In the above line of reasoning, we used Theorem 11-1 concerning an equivalent representation for the constant e.

It can also be shown (proof omitted) that $\sigma^2 = Var(X) = \lambda$.

Figure 47 depicts the graph of the Poisson distribution with $\lambda = 7$ and the number of events k along the horizontal axis. For example, the probability of 12 events in an interval is .0263. It is important to note that the tail is infinitely long with albeit very small probabilities as k becomes large.

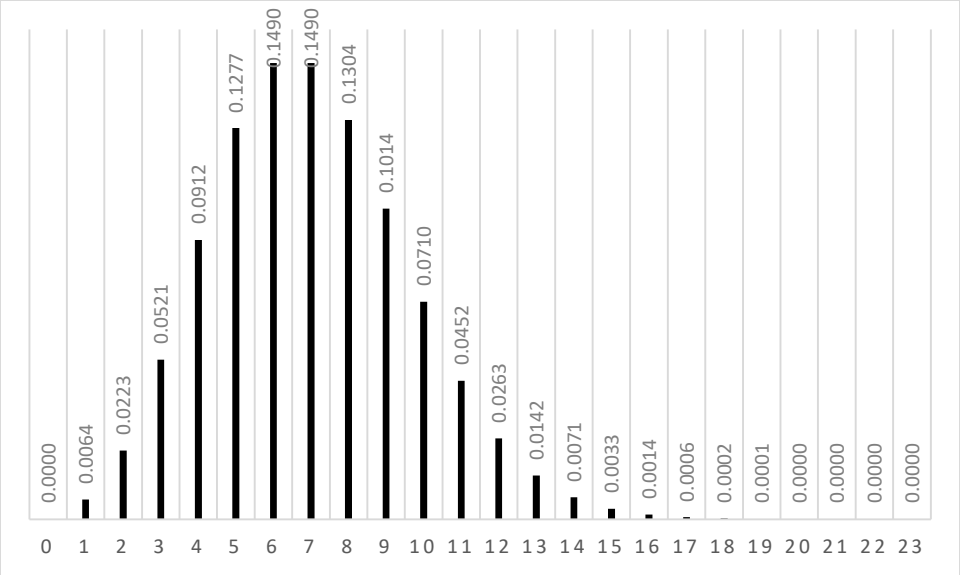


Figure 47. Graph of Poisson Distribution, $\lambda = 7$

12.9.4 Continuous Random Variables

12.9.4.1 Definitions and Some Examples

A continuous random variable is defined by a Probability Density Function (PDF). The probability that a continuous random variable X takes values in a given range (for example, between a and b) is computed by determining the area between the PDF and the x-axis from a to b. As described in the section on integral calculus, the area between a curve (in this case a PDF) and the x-axis between two points can be determined by computing the definite integral of the given function between the two points. It follows that for continuous PDFs, the probability at a single value is 0. So, we only talk about the probability for a range (i.e., interval) of values.

As an initial example, consider the continuous random variable X with PDF given by $f(x) = x/4$ over the interval $x = 1$ to 3. Note that $\int_1^3 x/4\, dx = \left[\frac{x^2}{8}\right]_1^3 = \frac{9}{8} - \frac{1}{8} = 1$ which should always be the case for a continuous PDF, i.e., the area under the PDF should be 1 over the domain of the PDF. The PDF and associated area under the PDF are shown in Figure 48.

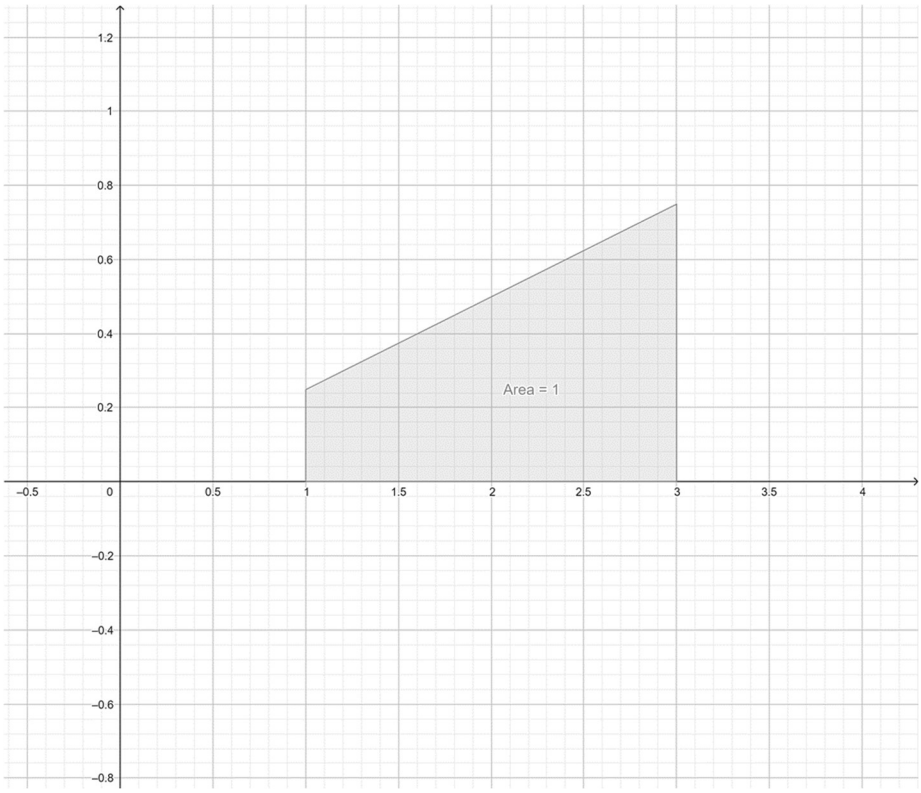


Figure 48. Area under f(x)=x/4 from x=1 to 3

If the probability that the random variable X takes on values between say 1 and 2 is desired, this can be computed as $\int_1^2 x/4\, dx = \left[\frac{x^2}{8}\right]_1^2 = \frac{4}{8} - \frac{1}{8} = \frac{3}{8}$.

The expected value for a continuous random variable is defined as $\int_a^b x\, f(x)\, dx$ where a and b are the lower and upper bounds for the random variable. As with discrete random variables, both $E(X)$ and μ are used to represent the expected value.

The variance of a continuous random variable is defined as $Var(X) = E((X-\mu)^2) = \int_a^b (x-\mu)^2\, f(x)\, dx$. As was the case for discrete random variables, it is also true that $Var(X) = E(X^2) - \mu^2$. Further, the standard deviation is defined as $\sigma = \sqrt{Var(X)}$.

For the random variable X with PDF equal to $f(x) = x/4$, the expected value is $\int_1^3 x(\frac{x}{4})\, dx = \int_1^3 \frac{x^2}{4} dx = \left[\frac{x^3}{12}\right]_1^3 = \frac{27}{12} - \frac{1}{12} = \frac{26}{12} \cong 2.167$. Next, we compute $E(X^2) = \int_1^3 x^2(\frac{x}{4})\, dx = \int_1^3 \frac{x^3}{4} dx = \left[\frac{x^4}{16}\right]_1^3 = \frac{81}{16} - \frac{1}{16} = 5$ and so, $Var(X) = 5 - \left(\frac{26}{12}\right)^2 = \frac{11}{36}$. Thus, $\sigma = \sqrt{\frac{11}{36}} \cong .553$.

It is interesting to note that the expected value of X is not the point that cuts-off .5 of the probability (area under the PDF). In fact, the .5 cut-off point is $\sqrt{5} = 2.236$. The expected value is (so to speak) the fulcrum point for the region under the PDF and not the .5 cut-off point. If the PDF is symmetric then the expected value and the .5 cut-off point will be equal.

As another example, take the random variable Y with PDF given by $f(x) = \frac{3}{x^4}$ from 1 to ∞. First, let's check to make sure the probability (area under the curve) adds to 1 over the stated domain for $f(x)$:

$$\int_1^\infty \frac{3}{x^4}dx = \left[-\frac{1}{x^3}\right]_1^\infty = 0 - (-1) = 1, \text{ noting that } \lim_{n\to\infty} -\frac{1}{x^3} = 0.$$

In terms expectation, variance and standard deviation, we have the following:

$$E(Y) = \int_1^\infty \frac{3x}{x^4}dx = \int_1^\infty \frac{3}{x^3}dx = \left[-\frac{3}{2}\frac{1}{x^2}\right]_1^\infty = 0 - \left(-\frac{3}{2}\right) = 1.5$$

$$E(Y^2) = \int_1^\infty \frac{3x^2}{x^4}dx = \int_1^\infty \frac{3}{x^2}dx = \left[-\frac{3}{x}\right]_1^\infty = 0 - (-3) = 3$$

$$Var(Y) = E(Y^2) - \mu^2 = 3 - (1.5)^2 = .75 \text{ and } \sigma = \sqrt{Var(Y)} \cong .866.$$

So, even though the PDF for random variable Y is infinitely long, the expected value is finite, as is the standard deviation.

12.9.4.2 Normal Distribution

The most important and commonly applied of the continuous probability distribution is the normal (or Gaussian) distribution. This is the so-called "bell curve" due to its shape.

The PDF for the normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

where x goes from $-\infty$ to ∞, μ is the mean and σ is the standard deviation. When the Euler constant e is raised to a complicated function (as is the case in the above equation), it is common to write $e^{g(x)}$ as exp[g(x)].

The corresponding CDF is given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]dt$$

There is no closed form for the above integral. So, the values of $F(x)$ need to be computed via a spreadsheet or other application. For example, the NORM.DIST(x, mean, standard_dev, TRUE) function in Microsoft Excel returns the cumulative probability from $-\infty$ to x for a normal distribution of given mean and standard deviation.

Figure 49 depicts three normal PDFs, all with $\mu = 0$. The curve with the highest peak has $\sigma = 1$, the curve with the second highest peak has $\sigma = 2$ and the flattest curve has $\sigma = 3$. As expected, the curve is more spread out as σ increases.

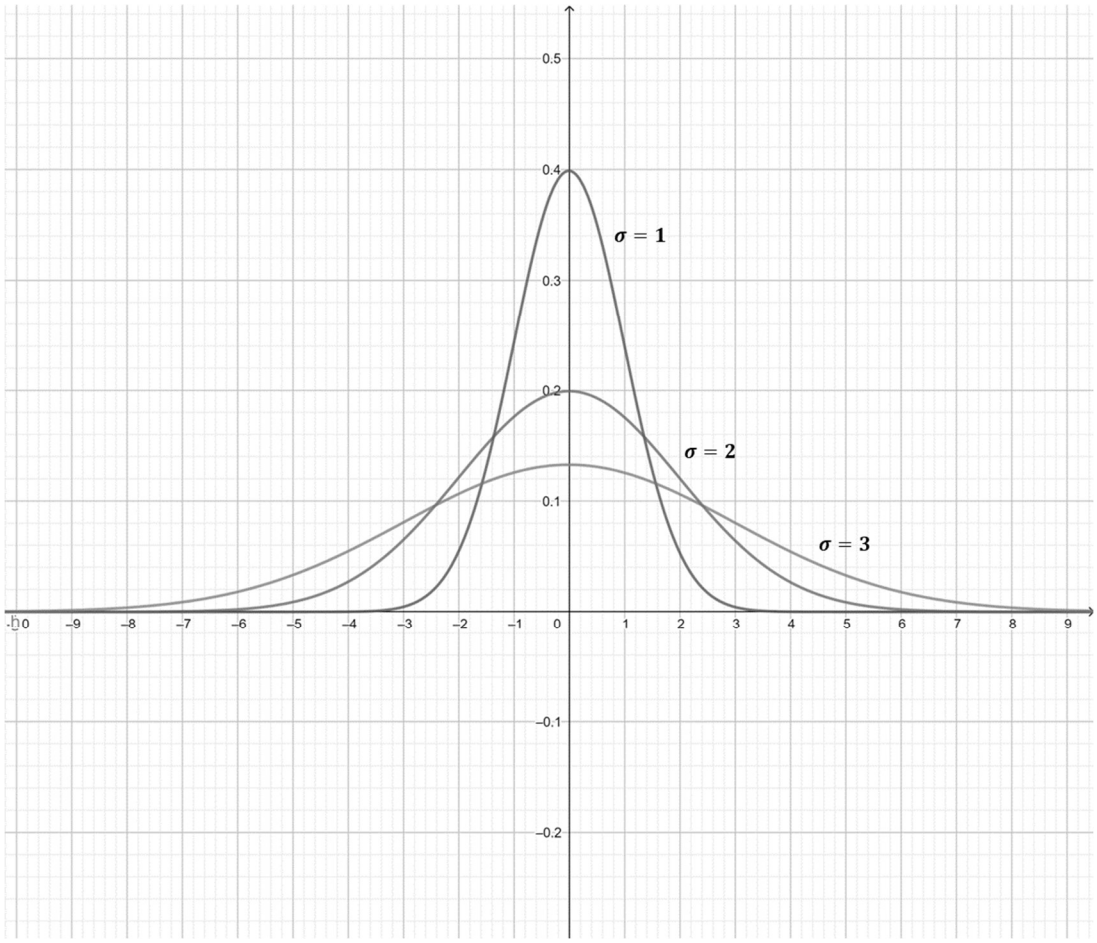


Figure 49. Normal Distributions

The **68–95–99.7 rule** (also known as the 3-sigma rule or empirical rule) is a mnemonic used to remember the percentage of values that lie within an interval around the mean in a normal distribution with a width of two, four and six standard deviations, respectively. In other words, the probability (area) under the PDF for a normal distribution in the intervals $(\mu - \sigma, \mu + \sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$ and $(\mu - 3\sigma, \mu + 3\sigma)$ are approximately .68, .95 and .997. For a normal distribution with $\mu = 0$ and $\sigma = 1$, Figure 50 shows the regions representing 68% (light gray area), 95% (light plus middle gray area) and 99.7% (light, middle and dark gray areas combined) of probability under the associated PDF.

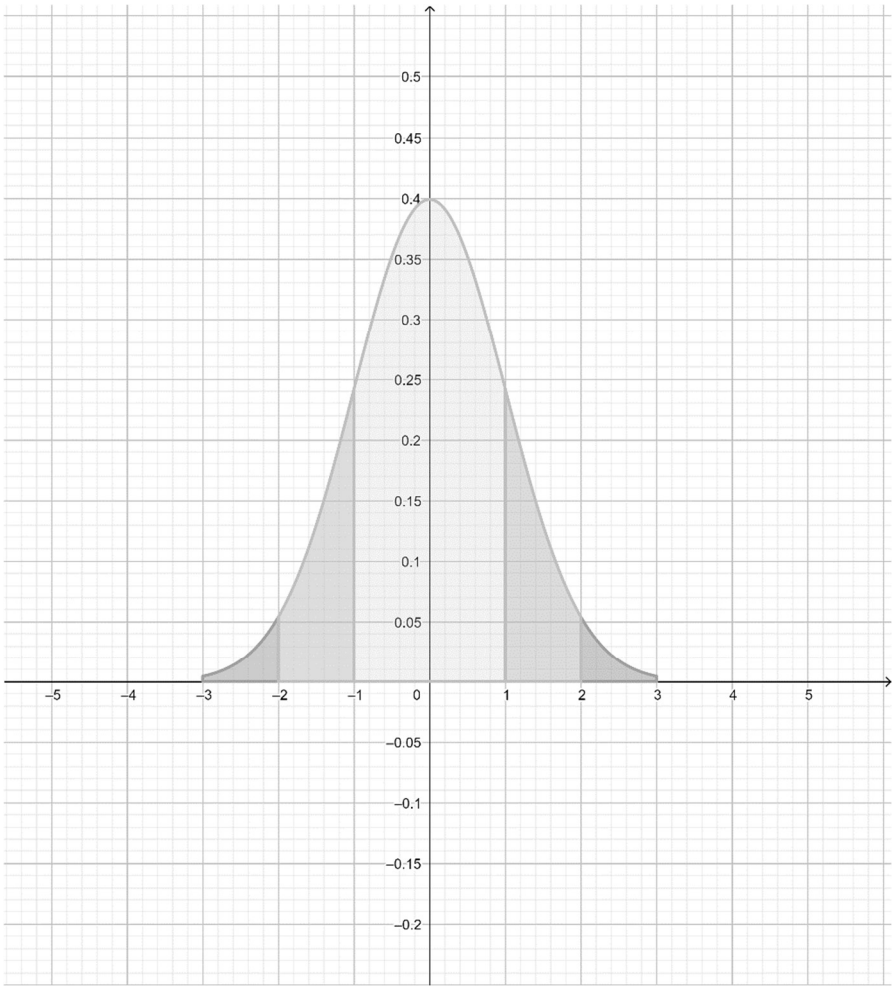


Figure 50. Illustration of the 68–95–99.7 rule

12.9.4.3 Exponential Distribution and the Memoryless Property

The PDF for a random variable X with the exponential distribution is given by the following formula, with $\lambda > 0$:

$$f(x) = \begin{cases} \lambda e^{-\lambda} & , x \geq 0 \\ 0, & x < 0 \end{cases}$$

The mean is given by $\mu = E(X) = \frac{1}{\lambda}$ and the variance is $\sigma^2 = 1/\lambda^2$.

Figure 51 shows three exponential distributions with $\lambda = 3$ (labeled f on the graph), $\lambda = 1.5$ (labeled g on the graph) and $\lambda = .75$ (labeled h on the graph).

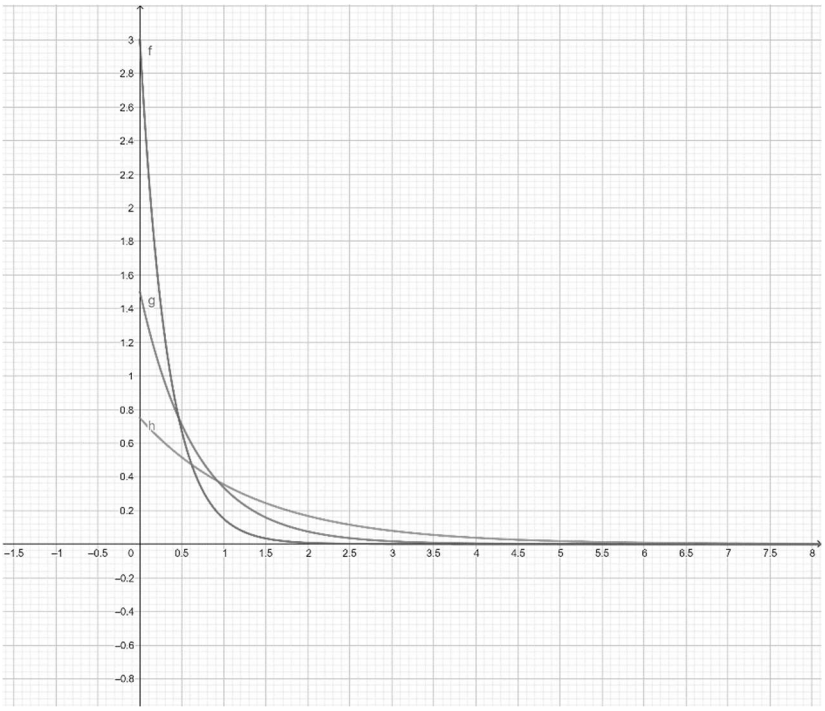


Figure 51. Exponential Distributions

Closely related to the exponential distribution is something called the memoryless (forgetfulness) property. If a probability distribution has the memoryless property, then the distribution is independent of its history. Given the memoryless property, the likelihood of something happening in the future has no relation to whether or not it has happened in the past, i.e., history has no impact on the future regarding the probability distribution.

More formally, a random variable X has the memoryless property if for every $s \geq 0, t \geq 0$, the following holds true: $P(X > s + t \mid x > s) = P(x > t)$. For example, if X represents the lifetime of hardware component in a computer, then probability a new component will last $t = 5$ or more years is the same as the probability the component will last another $t + s = 7$ or more years given that the component is already $s = 2$ or more years old.

In the following theorem, we have that the only memoryless continuous PDF is exponential. It should also be noted (but not proved here) that the only memoryless discrete PDF is the geometric distribution.

Theorem 12-10 A memoryless continuous PDF is necessarily exponential.

Proof: Given that a continuous random variable X satisfies the memoryless property, we have that $P(X > s + t \mid x > s) = P(X > t)$. By the definition of conditional probability and noting that $(X > s + t) \cap (X > s) = (X > s + t)$, the memoryless property can be rewritten as

$$\frac{P(X>s+t)}{P(X>s)} = P(X > t) \qquad \text{(Equation 1)}$$

Let $f(x) = P(X > x)$ and rewrite Equation 1 as

$$f(s + t) = f(s)f(t) \qquad \text{(Equation 2)}$$

Using Equation 2 and the principle of finite induction, we have that

$$f(kt) = f(t)^k \text{ for any positive integer } k \qquad \text{(Equation 3)}$$

Noting that $t = \frac{t}{k} + \frac{t}{k} + \cdots + \frac{t}{k}$ (k terms) for any positive integer k and using Equation 3, gives $f(t) = f\left(k\left(\frac{t}{k}\right)\right) = f\left(\frac{t}{k}\right)^k$ or equivalently,

$$f\left(\frac{t}{k}\right) = f(t)^{\frac{1}{k}} \qquad \text{(Equation 4)}$$

Now take any positive rational number $\frac{p}{q}$ (where p and q are positive integers). We have by the application of Equation 3 and then Equation 4 that

$$f\left(\frac{p}{q}t\right) = f\left(p\left(\frac{t}{q}\right)\right) = f\left(\frac{t}{q}\right)^p = f(t)^{\frac{p}{q}} \qquad \text{(Equation 5)}$$

Every positive real number is the limit of a sequence of rational numbers, e.g., $\lim_{n\to\infty} \frac{floor(xn)}{n} = x$ where x is any positive real number, n is a positive integer. Recall that the floor function is the greatest integer less than a given argument. Applying this concept to Equation 5, we have that

$$f(xt) = f(t)^x \text{ for any real number } x > 0 \qquad \text{(Equation 6).}$$

In Equation 6, let $t = 1$ to get $f(x) = f(1)^x = e^{\ln[f(1)]x}$ (recalling that e^z and $\ln z$ are inverse functions). So, $f(x)$ is of the form $e^{-\lambda}$ where $\lambda = -\ln[f(1)]$ ■

Regarding the limit in the above proof, it is helpful to compute some example values to visualize the convergence. For example, take $x = \pi = 3.14159\ldots$ which we know to be an irrational number. The following table shows how $\frac{floor(xn)}{n}$ starts to converge to x as we input large values of n.

n	10	100	1000	10000	100000	…
$floor(xn)/n$	$\frac{31}{10}$	$\frac{314}{100}$	$\frac{3141}{1000}$	$\frac{31415}{10000}$	$\frac{314159}{100000}$	…

12.9.5 Standardized Random Variables

It is possible to transform (or "standardize") a random variable X with expected value μ and variance σ^2 so that it has expected value 0 and variance 1.

We first need to state several theorems before explaining how to standardize a random variable.

Theorem 12-11 For random variables X and Y, and constant a, the following statements are true $E(aX) = aE(X)$ and $E(X + Y) = E(X) + E(Y)$.

A proof of this theorem can be found in the Proof Wiki article entitled *Linearity of Expectation Function* [53].

Theorem 12-12 If X is a random variable with finite variance, the $Var(aX) = a^2Var(X)$, for any constant a.

Proof: As noted previously, $Var(X) = E(X^2) - \mu^2$ and so, we have

$Var(aX) = E((aX)^2) - (E(aX))^2$

$= E(a^2X^2) - (aE(X))^2$ by Theorem 12-11

$= a^2E(X^2) - a^2E(X)^2$ by Theorem 12-11 and noting that X^2 is also a random variable

$= a^2Var(X)$■

If X and Y are independent random variables, the following theorem holds true.

Theorem 12-13 If X and Y are independent, $Var(aX + bY) = a^2Var(X) + b^2Var(Y)$.

The derivation for the $Var(aX + aY)$ can be found in the Proof Wiki article entitled *Variance of Linear Combination of Random Variables* [54]. The derivation in the Proof Wiki article does not assume X and Y to be independent and thus provides a more general formula, i.e.,

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$$

However, when X and Y are independent, $Cov(X, Y) = 0$. Covariance (or Cov for short) is a measure of the strength of the correlation between two random variables.

Now, back to the standardization of random variable X. Consider the random variable $Y = \frac{X-\mu}{\sigma}$.

Using Theorem 12-11 and noting that the expected value of a constant is the same constant, we have that $E(Y) = E\left(\frac{X}{\sigma}\right) - E\left(\frac{\mu}{\sigma}\right) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$.

Using Theorem 12-12 and substituting $\mu = 0$, we get $Var(Y) = Var\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2}Var(X) = \frac{\sigma^2}{\sigma^2} = 1$.

Thus, Y is a transformation of random variable X that has expected value 0 and variance 1.

12.9.6 Z-Score

A z-score (or sometimes z-value) is a measure of how far a given data point is from the mean of a distribution in terms of standard derivations. A z-score of 0 indicates that the data point is equal to the mean. A z-score of -1 indicates a value that is one standard deviation less than the mean. The process of converting a raw score into a standard score is called standardizing or normalizing [56].

For example, assume a student gets a result of 60 on a standardized test given to many thousands of other students. The result alone does not give much information as to how well the student did on the test compared to other students, even if you are told the possible scores range from 0 to 100. Let X be the random variable representing the results of all the students and assume the mean is $\mu = 50$ and the standard deviation is $\sigma = 3$. From this information, it is possible to compute how many standard deviations a given observed result x is from the mean, using the formula $z = \frac{x-\mu}{\sigma}$. For the problem at hand, we have $z = \frac{60-50}{3} = 3.33$ which is indeed an excellent result.

In general, $z = \frac{x-\mu}{\sigma}$ is the formula for computing the z-score, when both the mean and standard deviation are known. If the mean and standard deviation are not known, then we use the formula $z = \frac{x-\tilde{\mu}}{\tilde{\sigma}}$ where $\tilde{\mu}$ and $\tilde{\sigma}$ are the sample mean and standard deviation, respectively. Estimates for the sample mean and sample standard deviation are discussed in Section 13.2.

The definition of z-score does not assume a normal distribution and it is possible to compute a z-score for any distribution. In practice, however, the z-score is typically used under the assumption

of an underlying normal distribution. There are many online z-score calculators (just about all of which assume a normal distribution).

12.9.7 Law of Large Numbers

The **law of large numbers** is a proven theorem that describes the result of performing the same trial or observation a large number of times. The theorem states that the average of the results obtained from the trials or observations gets closer to the actual expected value as the number of trials or observations increases. The assumption is that each trial or observation adheres to the same probability distribution and that the probability distribution has a finite mean.

The following code (using the Python programming language) simulates rolls of a fair die. For each loop, the number of rolls is increased (up to 1000) and the mean of the rolls taken for each loop. As the number of rolls increases, the mean converges around the theoretical expected value of 3.5 (see Figure 52).

```
#Import numpy
import numpy as np
#Import pandas
import pandas as pd

results = []
for num_throws in range(1,1000):
   throws = np.random.randint(low=1,high=7, size=num_throws)
   mean_of_throws = throws.mean()
   results.append(mean_of_throws)
df = pd.DataFrame({ 'throws' : results})
from IPython.core.pylabtools import figsize
from matplotlib import pyplot as plt
figsize(11, 9)
df.plot(title='Law of Large Numbers – Roll of Die Simulation',color='b')
plt.xlabel("Number of throws in sample")
plt.ylabel("Average Of Sample")
```

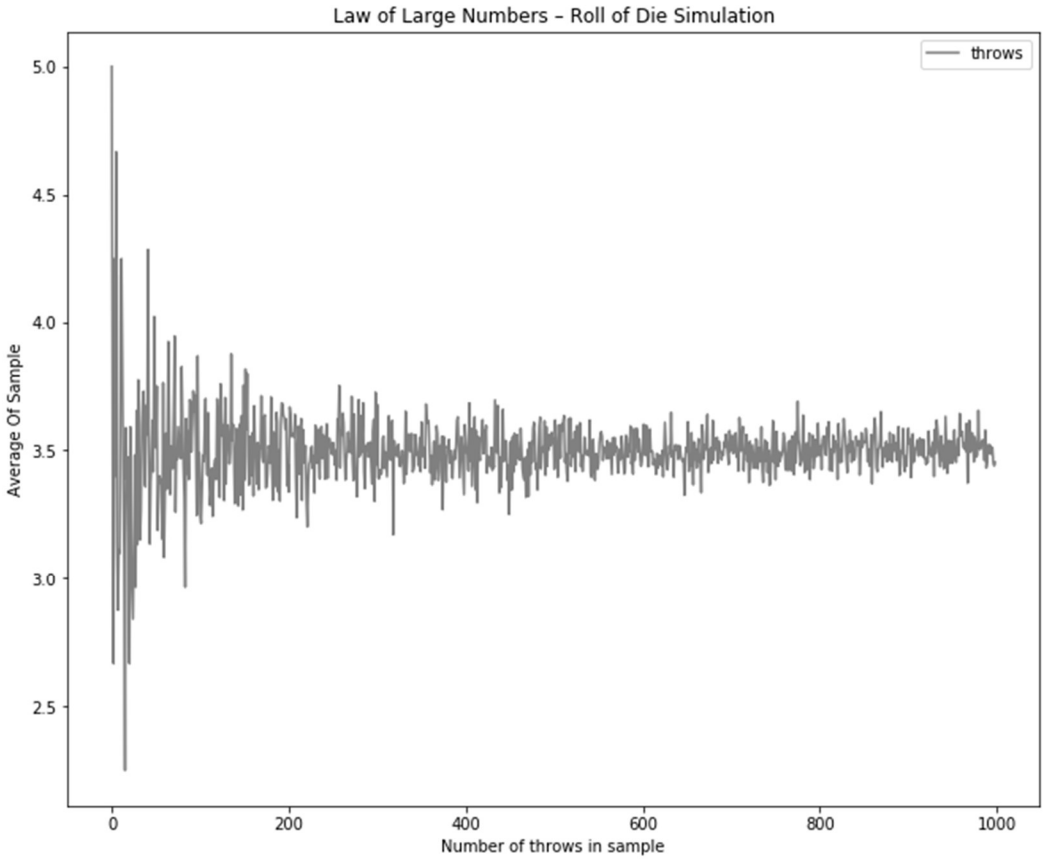


Figure 52. Simulated Roll of a Die

More formally, the law of large numbers is stated in the following theorem.

Theorem 12-14 (Law of Large Numbers) Let $X_1, X_2, \dots, X_n\}$ be a set of mutually independent and identically distributed random variables. If $\mu = E(X_i)$ for $i = 1,2, \dots, n$ exists (i.e., is finite), then for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|\bar{X}_n - \mu| > \varepsilon\} \to 0$$

where

$$\bar{X}_n = \frac{\mathrm{X_1} + \mathrm{X_2} + \cdots + \mathrm{X_n}}{n}$$

That is to say, the probability that the average of the samples differs from the expected value by more than any given value ϵ (however, small) approaches 0.

The above is known as the weak law of large numbers.

There is also something called the strong law of large numbers that states

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

The weak law guarantees that the probability of the difference between the sample mean and the actual expected value is more than some given small value ε goes to zero as the sample size goes to

infinity. The strong law, on the other hand, guarantees that events of the form "the sample mean is more than ε from the actual expected value" eventually stop happening.

An in depth discussion of the various types of convergence of random variables is well beyond the scope of this book. The interested reader will find a good discussion of the topic in Section 7 of the open access, peer-reviewed textbook entitled Introduction to Probability, Statistics and Random Processes [57].

The law of large numbers is discussed further in Section 18.2.

12.9.8 Central Limit Theorem

Related to the law of large numbers is the **central limit theorem** which states that the distribution of sample means (from a given distribution) approximates a normal distribution as the sample size increases, assuming that all samples are identical in size, and regardless of the shape of the original distribution (i.e., does not need to be of a bell-shape like the normal distribution).

The following Python code simulates 50,000 samples of size 1000 from each of four different distributions, i.e., binomial, exponential, geometric and Poisson. Figure 53 shows the result of each of the four simulations. In each case, the distribution of the sample means approximates a normal distribution as predicted by the central limit theorem.

```
#Import numpy
import numpy as np
#Import pandas
import pandas as pd

samples_all = []
samples_all_exp = []
samples_all_possion = []
samples_all_geometric = []
mu = .7
lam = 11
size = 1000
for number_in_sample in range(1,50000):
   samples = np.random.binomial(33, mu, size=size)
   samples_all.append(samples.mean())
   samples = np.random.exponential(scale=2.0,size=size)
   samples_all_exp.append(samples.mean())
   samples = np.random.geometric(p=.5, size=size)
   samples_all_geometric.append(samples.mean())
   samples = np.random.poisson (lam=lam, size=size)
   samples_all_possion.append(samples.mean())
df = pd.DataFrame({ 'binomial' : samples_all,
                    'poission' : samples_all_possion,
                    'geometric' : samples_all_geometric,
                   'exponential' : samples_all_exp})
from IPython.core.pylabtools import figsize
from matplotlib import pyplot as plt
figsize(17, 8)
fig, axes = plt.subplots(nrows=2, ncols=2)
df.binomial.hist(color='blue',ax=axes[0,0], alpha=0.9, bins=43)
df.exponential.hist(ax=axes[0,1],color='r',bins=43)
df.poission.hist(ax=axes[1,0],color='g',bins=43)
df.geometric.hist(ax=axes[1,1],color='black',bins=33)
axes[0,0].set_title('Binomial')
axes[0,1].set_title('Exponential')
axes[1,0].set_title('Poisson')
axes[1,1].set_title('Geometric')
```

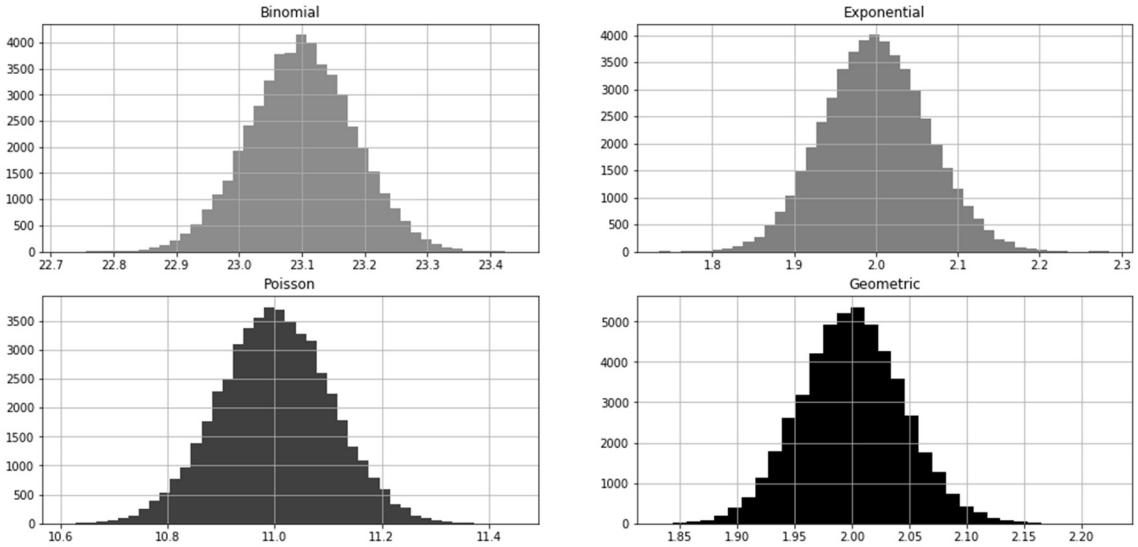


Figure 53. Simulation of Central Limit Theorem for Several Distribution Types

More formally, the central limit theorem is stated as follows:

Theorem 12-15 (Central Limit Theorem) Let $X_1, X_2, \dots, X_n\}$ be a set of identically distributed (i.e., same PDF) and independent random variables with expected value μ and finite variance σ^2. If $S_n = X_1 + X_2 + \cdots + X_n$, then the following is true

$$\lim_{n\to\infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\, dx$$

The integral on the right is the probability (between $x = a$ and $x = b$) of the normal distribution with expected value 0 and standard deviation 1.

By Theorem 12-11, random variable S_n has expected value $n\mu$ and by Theorem 12-13, $Var(S_n) = n\sigma^2$. Thus, $\frac{S_n - n}{\sigma\sqrt{n}}$ is a standardization of S_n with expected value 0 and variance 1.

Another way to state the central limit theorem is to say "as n approaches infinity, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ is approximated by a normal distribution with expected value 0 and variance 1."

The proofs of the law of large numbers and the central limit theorem go beyond the level of mathematical sophistication in this book. The interested reader can find proofs in the book "Introduction to Probability Theory" by Hoel, Port and Stone [58].

12.10 Exercises

1. If the odds against an event happening (e.g., a horse winning a race) is 50:1, what is the probability of the event happening? Convert the given odds to decimal odds and to moneyline odds. ***Hint**: Check you answer at https://www.gamblingsites.org/sports-betting/odds-converter/ or a similar odds conversion website.*

2. What's the probability that a randomly selected whole number between 1 and 150 (inclusive) has the property "not divisible by 2, 3 or 7"? ***Hint****: Use the result from Section 10.5.1.*
3. Three whole numbers between 1 and 150 are randomly selected. What is the probability that at least one of the numbers has the property "not divisible by 2, 3 or 7"? ***Hint****: This is the same as 1 minus the probability that all three numbers don't have the property "not divisible by 2, 3 or 7", i.e., they are divisible by 2, 3 or 7.*
4. Given sets $X = \{\boldsymbol{a}, b, c, d, \boldsymbol{e}, f\}$ and $Y = \{\boldsymbol{o}, p, q, r, s, t, \boldsymbol{u}\}$, and the task of randomly selecting one letter from each set.
    a. What is the probability of selecting d and q?
    b. What is the probability of selecting d or q?
    c. What is the probability of selecting exactly two vowels (in bold)?
    d. What is the probability of selecting at least one vowel? ***Hint****: This is 1 minus the probability of selecting no vowels.*
5. A team consists of seven men and five women. If two members of the team are randomly selected, what is the probability that both representatives are female? ***Hint****: Apply Theorem 12-6 to two events, i.e., selection of the 1st and 2nd female representative.*
6. Six women (their first initials are A, B, C, D, E and F) sit randomly in six chairs around a circular table. Two seating arrangements are considered to be the same if one can be rotated into the same position as the other. ***Hint****: There are a total of 5! seating arrangements.*
    a. What is the probability that A will have B to her immediate left, and C to her immediate right? ***Hint****: Given that B is to the left of A and C is to the right of A, determine the number of ways to arrange D, E and F.*
    b. What is the probability that A and B sit next to each other? ***Hint****: Count the number of ways A can be seated to the left of B, and the number of ways A can be seated to the right of B.*
7. An automated recognition system can with 95% accuracy identify a hand-written word. This is an early prototype and the possible words that can be recognized are from a list of 100 words. The system identifies a given hand-written word as a particular word in the given list (say the word "table" is identified). What is the probability the identification is correct? ***Hint****: Let A be the event the handwritten word is "table" and B be the event that the automated recognition systems identifies the word as "table." Assume the handwritten word is chosen randomly, i.e.,* $P(A) = \frac{1}{100}$*. Use Bayes' rule along with the law of total probability to compute* $P(A|B)$*.* ***Answer****:* $\frac{19}{118}$*.*
8. Two dice (faces numbered 1 to 6) are rolled and the sum is computed. A roll of 7 is considered a success and any other roll is considered a failure. What is the expected number of rolls before a success? ***Hint****: This is the geometric distribution with* $p = \frac{1}{6}$ *(probability of rolling a 7 with two dice). In other words, we are looking for the expected value of the geometric distribution with* $p = \frac{1}{6}$*.*

9. Verify the function $f(x) = \frac{1}{80}(x^3 - 7x)$ on the interval [3,5] is a valid PDF, i.e., area under the curve from 3 to 5 is equal to 1. Find the expected value and standard deviation for $f(x)$.
10. Show that $f(x) = 2/x^3$ is a valid PDF for random variable X defined on the interval 1 to ∞, i.e., the area under $f(x)$ for the stated interval is 1. Show that expected value of X is 2 but the variance of X is undefined. ***Hint**: $E(X^2)$ diverges, i.e., the associated integral goes to infinity.*
11. Show that $f(x) = 1/x^2$ is a valid PDF for random variable X defined on the interval 1 to ∞, but that neither the expected value nor the variance is defined, i.e., they go off to infinity.

13 Statistics

13.1 Overview

Statistics is what marries real life data to probability models so that predictions can be made. It also provides tools for understanding and analyzing data in a structured manner. There are two main classifications of statistics, i.e., descriptive statistics, which allow one to summarize data from a sample using measures such as the mean or standard deviation, and inferential statistics, which allow one to make hypotheses and then draw conclusions from data samples.

Recall, from Section 12.2, the mention of two basic approaches for the calculation of probabilities. Section 12 was mainly focused on the combinatorial approach which entails the counting of possibilities (basically, a non-empirical approach). This section focuses on an empirical approach where, for example, rather than saying each face of a die has a $\frac{1}{6}$ chance of being face-up after a roll, we actually roll a particular die and count the occurrence of each number and derive a probability distribution from the observations.

13.2 Descriptive Statistics

Descriptive statistics are used in everyday topics such as stock market averages, various sports statistics such as baseball batting averages, viewer-based movie ratings and product ratings on online retail websites.

For a given statistic, such as a baseball batting average, there are several basic measures used to describe the data, i.e., central tendency measures such as the mean, dispersion from the mean (how spread-out are the data points) and the shape of the distribution.

In the case of potentially related statistics (e.g., high school grade point average and college grade point average), statisticians typically look for measures of correlation or lack thereof.

13.2.1 Central Tendency

Consider the set of (hypothetical) ratings of a given movie in Table 36, where each viewer can give a score from 1 to 10. The movie ratings are in the second row. The first row is an index that allows for reference to a specific item in the list.

Table 36. Movie Rating Example

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
3	3	4	4	4	5	5	5	5	6	6	6	6	7	7	7	8	8	9

The sample **mean** $\tilde{\mu}$ is the sum of the numbers divided by the number of data elements. For the example at hand, $\tilde{\mu} = \frac{108}{19} = 5.68$.

Table 37 shows the frequency for each rating and the associated proportion, e.g., two people gave the movie a rating of 3, with associated proportion $\frac{2}{19}$ of the total number of ratings. This gives us another way to compute the mean in a way analogous to the expected value, i.e., $\tilde{\mu} = \sum_{i=1}^{n} p_i x_i$

where p_i is the proportion (probability estimate) of the data points with value x_i. For the example at hand, we have $\tilde{\mu} = 3\left(\frac{2}{19}\right) + 4\left(\frac{3}{10}\right) + 5\left(\frac{4}{19}\right) + 6\left(\frac{4}{19}\right) + 7\left(\frac{3}{19}\right) + 8\left(\frac{2}{19}\right) + 9\left(\frac{1}{19}\right) = \frac{108}{19} = 5.68.$

Table 37. Movie Ratings – Frequencies and Proportions

Viewer Rating	3	4	5	6	7	8	9
Frequency	2	3	4	4	3	2	1
Proportion	$\frac{2}{19}$	$\frac{3}{19}$	$\frac{4}{19}$	$\frac{4}{19}$	$\frac{3}{19}$	$\frac{2}{19}$	$\frac{1}{19}$

If the data points are arranged in ascending order, the **median** is the half-way point. Consider Table 38 which shows the occurrence of various salaries (in $1000) for the 101 employees at a particular company. The data is a bit skewed in that the four corporate officers make much more than the rank and file. The sample mean is

$$\tilde{\mu} = 50\left(\frac{12}{101}\right) + 75\left(\frac{60}{101}\right) + 100\left(\frac{15}{101}\right) + 125\left(\frac{10}{101}\right) + 10000\left(\frac{4}{101}\right) = \frac{47850}{101} \cong 473.7624$$

or about $473,762 per year, which is misleading. In such cases, the median may give a better summary of the data. For the example at hand, the median is $75,000 since there are 50 data points below and above this salary. (The example may be slightly confusing since there are multiple appearances of the data point $75,000.) Another approach would be to leave the salaries of the corporate officers out of the calculation and just compute the mean salary for the rank and file.

Table 38. Salary Frequencies

Salary in $1000	50	75	100	125	10000
Frequency	12	60	15	10	4
Proportion	$\frac{12}{101}$	$\frac{60}{101}$	$\frac{15}{101}$	$\frac{10}{101}$	$\frac{4}{101}$

While the median divides the data in half, **quartiles** divided the data into quarters. This approach can be extended to q equal subdivisions of the ordered data, with each subdivision known as a q-quantile. This is particularly useful with large sets of data.

Yet another approach is to use **percentiles**, i.e., a measure indicating the value below which a given percentage of data points falls. Percentiles are sometimes used in standardized tests. For example, a student may be told his or her test score is at the 75th percentile, meaning 75% of the people who took the test had lower scores and 25% had higher scores. This does not mean the student got a score of 75 out of a 100 on the test. For example, the test might have been very hard and a score of 50 out of 100 might equate to the 75th percentile.

Another measure is the **mode**, i.e., the data value that occurs most often. Table 39 shows the number of failures for a given type of automobile component. The total number of components is 251. There are many failures in the first month (perhaps a manufacturing issue), very few failures for months 2-12 and then a lot of failures in month 13 (perhaps the part either fails or is required to be replaced after 12 months). The mean is $\frac{1573}{251} \cong 7.5$ months which is not helpful in summarizing

the data since there are no failures at all in month 7 and very few failures in the surrounding months. The median is 12, i.e., 125 values fall before month 12 and 125 fall after the first data point in month 12. This is also misleading and neglects the concentration of failures in month 1. The mode is 13 but that's just half the story. A better approach would be to claim the distribution is bimodal with peaks at month 1 and 13.

Table 39. Frequency of Component Failures per Month

Component Failures in Month x	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	100	5	4	3	2	1	0	1	2	3	4	5	121
Proportion	$\frac{100}{251}$	$\frac{5}{251}$	$\frac{4}{251}$	$\frac{3}{251}$	$\frac{2}{251}$	$\frac{1}{251}$	0	$\frac{1}{251}$	$\frac{2}{251}$	$\frac{3}{251}$	$\frac{4}{251}$	$\frac{5}{251}$	$\frac{121}{251}$

13.2.2 Dispersion

The dispersion of a sample distribution can be measured by the standard deviation of the sample.

If the mean μ of the total population is known, then the standard deviation for the set of sample points $\{x_1, x_2, \dots, x_n\}$ is given by

$$\tilde{\sigma} = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots (x_n - \mu)^2}{n}}$$

This formula follows directly from the definition of variance for a discrete random variable (see Section 12.9.3.1) with the probability of each observation assumed to be equal, i.e., probability $\frac{1}{n}$.

If the mean is not known, then the sample mean is used, i.e.,

$$\tilde{\mu} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The standard deviation, in this case, is usually determined by the following formula

$$\hat{\sigma} = \sqrt{\frac{(x_1 - \tilde{\mu})^2 + (x_2 - \tilde{\mu})^2 + \cdots (x_n - \tilde{\mu})^2}{n - 1}}$$

Division by $n - 1$ rather than n is intended to partially correct for the estimation error caused by using the sample mean. One can show that the expected value of $\hat{\sigma}$ is equal to the actual standard deviation, i.e., $\hat{\sigma}$ is an unbiased estimator of σ. This modification to the formula for standard deviation is known as Bessel's correction and is explained further in the Wikipedia article on this topic [62].

One last point concerning dispersion: it can be argued that the concepts of quartile, q-quantile and percentile are measures of dispersion.

13.2.3 Shape of a Probability Distribution Function

The shape of a distribution (as revealed by a histogram) may also reveal some insights into the nature of the distribution. A **histogram** provides a visual presentation of numerical data by indicating the number of data points that lie within a range of values.

The movie rating example in the previous section can be represented with the histogram shown in Figure 54. The various ratings are listed on the horizontal axis and the number of people who gave a particular rating is represented by the height of the associated bar, e.g., 3 people gave a rating of 4 to the movie.

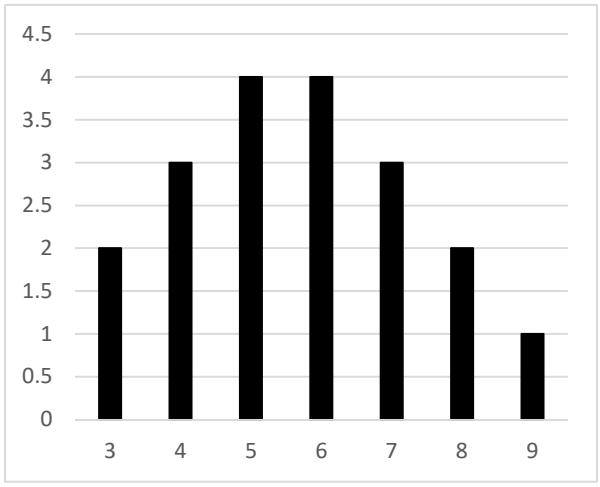


Figure 54. Histogram for Movie Rating Example

The histogram for the component failure example is depicted in Figure 55.

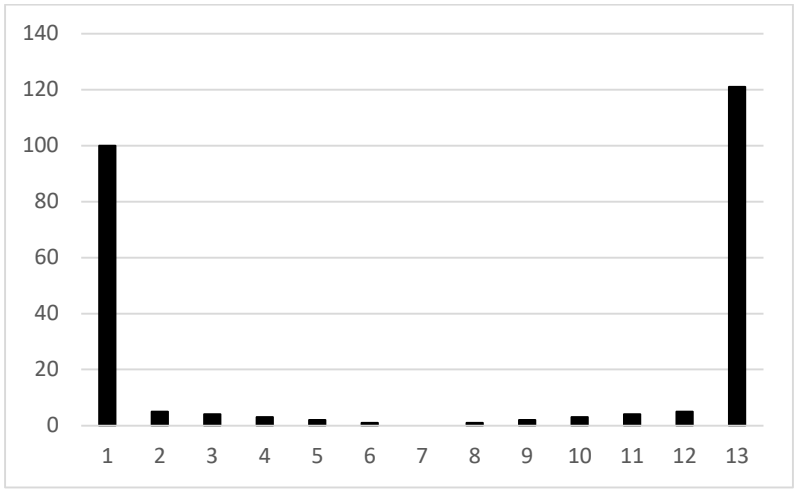


Figure 55. Histogram for Component Failure Example

In the previous examples, each bar represented a single value for a random variable, where the height of each bar is proportional to the frequency of the associated value. It is also possible to group a range of values. In this case, it is important the value ranges are of the same size. The grouping of values approach was used in the four histograms shown in Figure 53. In the associated

Python code, the *bins* parameter was used to indicate how many bars should be used in the histogram, with each bar representing a range of values along the horizontal axis.

13.3 Inferential Statistics

Inferential statistics entails procedures that allow one to infer patterns about a population based on study of a sample of the population. Practitioners of statistics use inferential statistics to examine the relationships between variables within a sample, and to then generalize or predict how those variables relate in the overall population.

Since it is typically impractical or impossible to examine all members of a given population, statisticians choose a representative subset of the population (a random sample) and attempt to draw conclusions about the entire population based on a study of the sample (or set of samples). The concept of drawing inferences about a population from a sample (taken from that population) is depicted in Figure 56.

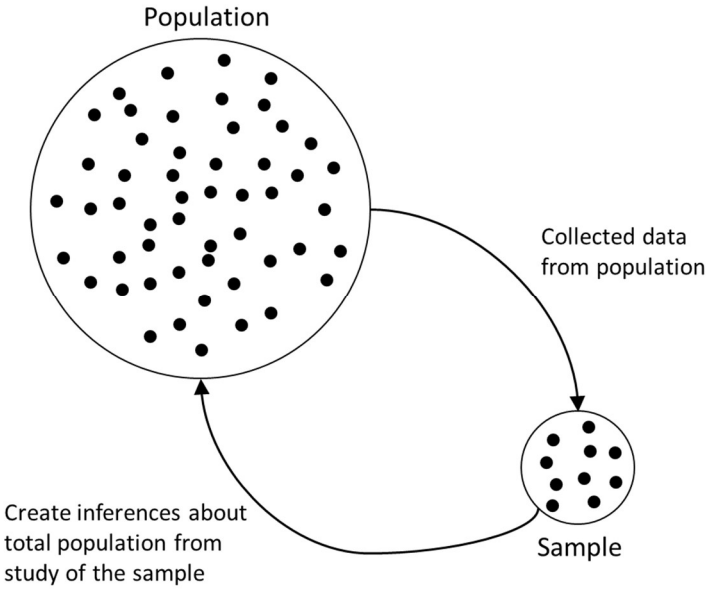


Figure 56. Sampling from a Population

There are two main divisions of inferential statistics:

- Confidence intervals which give a range of values for an unknown parameter of the population by measuring a statistical sample
- Hypothesis testing where one makes a claim about the population by analyzing a statistical sample.

13.3.1 Point Estimators

Point estimation is the process of finding an approximate value of some parameter (e.g., mean, variance or proportion) using random samples from a given population. Point estimators provide a single value approximation for a given parameter whereas confidence intervals (discussed in the next section) give a range of possible values for a given parameter.

It is desirable for a point estimate to have the following characteristics:

- Consistent – the larger the sample size, the more accurate the estimate
- Unbiased – the expected value of the point estimator equals the parameter being estimated. For example, if $X_1, X_2, \ldots, X_n$ are random variables that represent the results of random samples from a population with mean μ, then $\tilde{\mu} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ is an unbiased estimator of μ since $E(\tilde{\mu}) = E[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)] = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}(n\mu) = \mu$ by repeated application of Theorem 12-11 and noting that $E(X_i) = \mu$ for $i = 1,2, \ldots, n$.
- Most efficient or best unbiased – out of all the consistent and unbiased estimators, choose the one having the smallest variance. In other words, the estimator that varies least from sample to sample.

13.3.2 Confidence Intervals

A **confidence interval** is an estimate of the value for a parameter, computed from observed data. The estimate is stated in terms of a range of plausible values (i.e., an interval) for the unknown parameter. Associated with a confidence interval is a **confidence level** (a probability) that the given interval covers the unknown parameter value. In this document, the confidence level is represented by the Greek letter γ. Confidence intervals are often represented in the form $(x - z, x + z)$ where x is a point estimate for a parameter whose value is unknown and z is constant. It is also possible to have confidence intervals that are not symmetric about a particular value but such cases are not covered here.

From the Wikipedia article on Confidence Intervals [63]:

> More strictly speaking, the confidence level represents the frequency (i.e., the proportion) of possible confidence intervals that contain the true value of the unknown population parameter. In other words, if confidence intervals are constructed using a given confidence level from an infinite number of independent sample statistics, the proportion of those intervals that contain the true value of the parameter will be equal to the confidence level. For example, if the confidence level is 90% then in a hypothetical indefinite data collection, in 90% of the samples the interval estimate will contain the population parameter.

When forming a confidence interval for a population mean, often one of two different probability distributions are used to model confidence intervals, i.e., the normal distribution (which we have already seen) and something called Student's t-distribution (or just the t-distribution for short). The t-distribution is bell-shaped like the normal curve but it has more area in the tails of the distribution. The following is a typical decision path concerning when to use the normal distribution versus when to use the t-distribution:

- If the variance is known, use the normal distribution
- If the variance is not known:
    - If the sample size is large (greater than 30), use the normal distribution (referred to as a **z-test**)

- If the sample size is small (30 or less), then use the t-distribution (referred to as a **t-test**). In this case, the sample variance is used in place of the population variance.

As the sample size becomes large, the t-distribution approximates the normal distribution. So, for large sample sizes, one could use either distribution with similar results.

13.3.2.1 Using Normal Distribution to Compute Confidence Interval for the Mean

Assume that X is a random variable with known variance σ^2 and unknown mean μ. Let $X_1, X_2, \ldots, X_n$ be a random sample from X and let $\tilde{\mu}$ be the sample mean. From the central limit theorem (Theorem 12-15) along with a rearrangement (i.e., divide numerator and denominator by n), we have that $\frac{\tilde{\mu}-\mu}{\sigma/\sqrt{n}}$ is approximately a normal distribution with expected value 0 and standard deviation 1 as n becomes large, i.e., as the sample size becomes large.

Next, consider $P\left(-z \leq \frac{\tilde{\mu}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = \gamma$ where γ is the confidence level. This can be rewritten (with some algebraic manipulation) as $P\left(\tilde{\mu} - \frac{z\sigma}{\sqrt{n}} \leq \mu \leq \tilde{\mu} + \frac{z\sigma}{\sqrt{n}}\right) = \gamma$ which should be interpreted as the probability that the random interval $\left(\tilde{\mu} - \frac{z\sigma}{\sqrt{n}}, \tilde{\mu} - \frac{z\sigma}{\sqrt{n}}\right)$ contains μ is equal to γ.

The steps for computing a confidence interval for confidence level γ are as follows:

1. Decide on the confidence level γ for the parameter under study. Make sure the assumptions for using the z-score are met, i.e., variance is known and sample size is greater than 30.
2. Obtain the required sample data and compute the sample mean $\tilde{\mu}$.
3. Determine the value of z which depends on the desired confidence level. Some of the more common confidence levels and associated z-values are listed in the following table:

Table 40. Confidence Levels and Associated z-value

Confidence Level (γ)	z-value
.80	1.28
.90	1.645
.95	1.96
.98	2.33
.99	2.58

4. Plug the values into the interval formula to get the desired confidence interval:

$$\left(\tilde{\mu} - \frac{z\sigma}{\sqrt{n}}, \tilde{\mu} + \frac{z\sigma}{\sqrt{n}}\right)$$

As an example, let X be a random variable that represents the charge duration for a type of a rechargeable battery. Assume that 100 batteries are tested and the average charge duration is $\tilde{\mu} = 300$ hours. Further, we are given that the actual standard deviation is $\sigma = 3$ hours. The task is to determine a 95% confidence interval for the actual mean μ. Since $\frac{\tilde{\mu}-\mu}{\sigma/\sqrt{n}}$ approximates a normal distribution with expected value 0 and standard deviation 1, we can use a calculator or table such as Table 40 to look up the value of z that gives a probability of .95 which happens to be $z = 1.96$. Using the expression derived above, we have

$$\left(\tilde{\mu} - \frac{z\sigma}{\sqrt{n}}, \tilde{\mu} + \frac{z\sigma}{\sqrt{n}}\right) = \left(300 - \frac{3}{10}(1.96), 300 + \frac{3}{10}(1.96)\right) = (299.412, 300.588)$$

So, $(299.412, 300.588)$ is a 95% confidence interval for μ, meaning the interval $(299.412, 300.588)$ contains μ with probability .95. Notice that as the sample size increases (with the confidence level remaining the same), the confidence interval becomes smaller (which makes sense intuitively).

13.3.2.2 Using Student's T-distribution to Compute Confidence Interval for the Mean

Assume that X is a random variable with unknown variance σ^2 and unknown mean μ. Let $X_1, X_2, \ldots, X_n$ be a random sample from X and let $\tilde{\mu}$ be the sample mean. Further, assume the sample size is small (less than 30) and so, the central limit theorem does not apply here. In such cases, the variance needs to be estimated using the sample data.

If we use Bessel's correction to estimate the variance, i.e., $\hat{\sigma} = \sqrt{\frac{(x_1-\tilde{\mu})^2+(x_2-\tilde{\mu})^2+\cdots(x_n-\tilde{\mu})^2}{n-1}}$, it can be proved that $\frac{\tilde{\mu}-\mu}{\hat{\sigma}/\sqrt{n}}$ follows a t-distribution. The proof is complex and omitted here.

As with the previous case (but now using the t-distribution instead of the normal distribution), we seek $P\left(-t \leq \frac{\tilde{\mu}-\mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t\right) = \gamma$ where γ is the confidence level. This should be interpreted as the probability that the random interval $\left(\tilde{\mu} - \frac{t\hat{\sigma}}{\sqrt{n}}, \tilde{\mu} + \frac{t\hat{\sigma}}{\sqrt{n}}\right)$ contains μ is equal to γ.

The steps for computing a confidence interval for confidence level γ are as follows:

1. Decide on the confidence level γ for the parameter under study.
2. Obtain the required sample data and compute the sample mean $\tilde{\mu}$ and estimate for the standard deviation $\hat{\sigma}$.
3. Determine the value of t which depends on the desired confidence level and on the degrees of freedom (the sample size minus one). A table of t-values for given confidence levels and degrees of freedom is shown in Table 41. The degrees of freedom are listed in the extreme left column. The table is taken from the Wikipedia article on the Student's t-distribution [65].
4. Plug the values into the interval formula to get the desired confidence interval: $\left(\tilde{\mu} - \frac{t\hat{\sigma}}{\sqrt{n}}, \tilde{\mu} + \frac{t\hat{\sigma}}{\sqrt{n}}\right)$.

Table 41. Table of t-values for one-side and two-sided confidence intervals

One-sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two-sided	**50%**	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.5%**	**99.8%**	**99.9%**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.080	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

We now return to the movie rating example from Section 13.2.1 and compute a 90% confidence interval for the average rating. We've already computed the sample mean and standard deviation, i.e., $\tilde{\mu} = 5.68$ and $\hat{\sigma} = 1.7$. The $df = n - 1 = 18$ (df stands for "degrees of freedom). For $df = 18$

and $\gamma = .90$, the associated t-value is $t = 1.734$ (in Table 41, see the value shown in red). Substituting into the interval formula in Step #4 above, we get

$$\left(\tilde{\mu} - \frac{t\hat{\sigma}}{\sqrt{n}}, \tilde{\mu} + \frac{t\hat{\sigma}}{\sqrt{n}}\right) = (5.68 - .676, 5.68 + .676) = (5.004, 6.356).$$

Thus, the interval $(5.004, 6.356)$ contains μ with probability .9.

13.3.2.3 Confidence Intervals for Proportions

A binomial proportion confidence interval is an estimated interval for the probability of success p calculated from the outcome of a series of n success–failure experiments (Bernoulli trials).

For example, let's say that we want to determine the number of people who will vote for Candidate A or Candidate B. Each person polled constitutes an experiment. Assume n people are polled and n_A favor Candidate A and n_B favor Candidate B. Then $\tilde{p} = \frac{n_A}{n_A + n_B}$ can be used as an estimate for the proportion of votes (from the total voting population) who will vote for Candidate A.

One approach for determining a binomial confidence interval uses the normal distribution to approximate the distribution of the mean of n independent Bernoulli random variables $X_1, X_2, \ldots, X_n$ where

- $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{np}{n} = p$, noting that the expected value of the Bernoulli distribution is just p (the probability of a success for a given trial)
- $\text{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$, noting that the variance of the Bernoulli distribution is $p(p-1)$.

Use $\tilde{p} = \frac{n_S}{n}$ as an estimate for p, where n_S is the number of successes in n Bernoulli trials.

An estimate for the standard deviation of $\frac{1}{n}\sum_{i=1}^{n} X_i$ is given by $\bar{\sigma} = \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}$.

As noted previously, we have from the central limit theorem that $\frac{\tilde{p}-p}{\bar{\sigma}}$ is approximately a normal distribution with expected value 0 and standard deviation 1 as n becomes large. Because the normal approximation is not accurate for small values of n, a good rule of thumb is to use the normal approximation only if $np(1-p) > 10$.

We seek $P\left(-z \leq \frac{\tilde{p}-p}{\bar{\sigma}} \leq z\right) = \gamma$ where γ is the confidence level. This should be interpreted as the probability that the random interval $(\tilde{p} - z\bar{\sigma}, \tilde{p} + \overline{z\sigma}) = \left(\tilde{p} - z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right)$ contains p is equal to γ.

Back to the election example, assume that 100 people are polled to see if they prefer Candidate A or B. The result is that 52 prefer A and 48 prefer B. So, $\tilde{p} = .52$ where we consider a preference for Candidate A to be a success. If we want a 95% confidence interval, then Table 40 indicates that we should use $z = 1.96$. Making substitutions into the interval formula above, we get

$$\left(\tilde{p} - z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}, \tilde{p} + z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}}\right) = (.52 - .098, .52 + .098) = (.422, .618)$$

which is a very wide confidence interval compared to typical election polls.

How big do we need to make n, to get the interval length under say 6% (3% on either side of the mean) while keeping at a 95% confidence interval?

We want to solve for n such that $z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} < .03$ which is equivalent to $n > \left(\frac{1}{.03^2}\right)\tilde{p}(1-\tilde{p})z^2 \cong 1066$. If we wanted to get within 1% on either side of the mean, it would take a sample size of about 9589.

This is why pollsters reporting that, for example, a given candidate has a 52% to 48% edge over another candidate often add the caveat that the difference is within the statistical error of the survey.

13.3.3 Hypothesis Testing

13.3.3.1 Overview

The process of deciding between the possible conclusions of an experiment, with the aid of probability theory, is known as hypothesis testing. In particular, hypothesis testing is used to infer whether a claim about a population is valid based on an analysis of sample data taken from a larger population. For example, if a drug company wants to decide if a vaccine is 95% effective (i.e., induces antibodies, in the recipient, to the given virus), a formal hypothesis test would be performed (this included the collection of sample data from a larger population). Statistical theory is used to decide on whether to accept the hypothesis or not.

In its simplest form, hypothesis testing entails a decision between two competing propositions called the **null hypothesis** (denoted as H_0) and an **alternative hypothesis** (denoted H_1). The approach taken is to assume the null hypothesis is true unless proven otherwise by overwhelming evidence (sort of like a defendant in a criminal trial, i.e., innocent until proven guilty beyond a reasonable doubt). As another example, consider a personal care product company that wants to show their toothpaste (Brand A) prevents cavities better than another company whose toothpaste is Brand B. The null hypothesis would state that the two brands are the same (e.g., same average number of cavities in a given sample size) and the hope is to find a result that makes the null hypothesis highly unlikely and thus leads to the acceptance of the alternative hypothesis, i.e., Brand A leads to significantly fewer cavities than Brand B.

A **test statistic** is an expression (a quantity derived from the sample such as the mean or variance) used to test an hypothesis. An hypothesis test is usually specified in terms of a test statistic. The test statistic is used to summarize sample data, typically reducing the data to one value that can be used to perform the hypothesis test. In general, a test statistic is determined in such a way as to quantify, relative to sample data, behaviors that distinguish the null hypothesis from the alternative hypothesis. There are many commonly used test statistics, some of which are listed in the Wikipedia article entitled “Test statistic” [66].

Two types of errors can occur when attempting to distinguish between the null hypothesis and the alternative hypothesis.

- The first type of error (known as a **Type 1 error**) occurs when the null hypothesis is wrongly rejected. The probability of a Type 1 error is represented by the Greek letter α and is referred to as the **level of significance**. So, $\alpha = P(H_0 \text{ is rejected} | H_0 \text{ is true})$.
- The second type of error (known as a **Type 2 error**) occurs when the null hypothesis is wrongly **not** rejected. The probability of a Type 2 error, i.e., $P(H_0 \text{ is not rejected} | H_1 \text{ is true})$, is denoted by the Greek letter β. The **power of an hypothesis test** is defined to be the probability that the test rejects the null hypothesis when the alternative hypothesis is true. So, the power of an hypothesis test is $1 - \beta$.

The two types of errors are summarized in Table 42.

Table 42. Type I and Type II Errors Regarding Hypothesis Testing

	$\boldsymbol{H_0}$ **is not rejected**	$\boldsymbol{H_0}$ **is rejected**
$\boldsymbol{H_0}$ **is true**	Correct Decision (probability $1 - \alpha$)	Type I error (probability α)
$\boldsymbol{H_1}$ **is true**	Type II error (probability β)	Correct Decision (probability $1 - \beta$)

Since α is typically chosen to be small (.05 or less), the decision to retain (not reject) H_0 implies not that H_0 is probably true, but only that H_0 could be true, whereas the decision to reject H_0 implies that H_0 is probably false (and that H_1 is probably true).

For a fixed sample size, it is not possible to control both types of errors, i.e., as one type of error goes down, the other type of error goes up.

13.3.3.2 Steps in an Hypothesis Test

Step 1: Set up hypotheses and select the level of significance α.

H_0 (null hypothesis): $x = c$ (e.g., mean of some distribution is equal to c), where c is a constant

H_1 (alternative hypothesis): $x \neq c$ (or $x < c$ or $x > c$)

As suggested above, the alternative hypothesis can take several forms, i.e.,

- $H_1: x > c$, where a higher value is hypothesized. This type of test is called an upper-tailed (or right-hand) hypothesis test (see the upper-right graph in Figure 57). If the estimate for parameter *x* (based on a sample from the population) is much greater than *c*, then we accept the alternative hypothesis.
- $H_1: x < c$, where a lower value is hypothesized. This is called a lower-tailed (or left-hand) hypotheses test (see the upper-left graph in Figure 57). If the estimate for parameter *x* (based on a sample from the population) is much smaller than *c*, then we accept the alternative hypothesis.
- $H_1: x \neq c$, where a difference is hypothesized. This is called a two-tailed test (see the graph at the bottom of Figure 57). In this case, the level of significance is split in half for each tail. If the estimate for parameter *x* (based on a sample from the population) is much smaller or much greater than *c*, then we accept the alternative hypothesis.

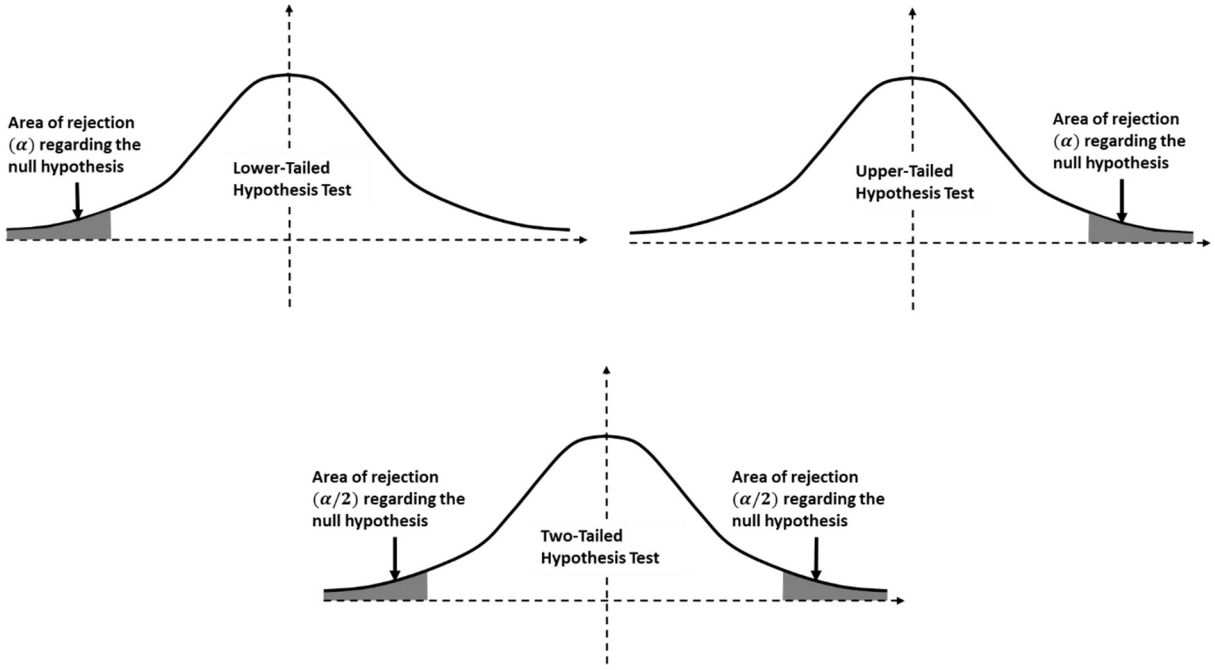


Figure 57. One and Two Tailed Hypothesis Tests

Step 2: Select the appropriate test statistic, i.e., a single number that summarizes the sample information. The test statistic is used to determine if the variable in the hypothesis (i.e., x in Step 1) is in the rejection region or not. For example, if the null hypothesis claimed the average temperature in an office was $72^{\circ}F$, the test statistic might be the mean (or median) temperature of hourly measurements taken over 3 days.

Step 3: Collect the required data to compute the test statistic, and then compute the test statistic.

Step 4: Arrive at a conclusion about whether to reject or not reject the null hypothesis.

Step 5: Possibly iterate the previous steps. Using the average temperature example above. Let's say that the claim of $72^{\circ}F$ was badly off the mark (test statistic resulted in a mean temperature of $65^{\circ}F$) and the null hypothesis was rejected. We could state a new null hypothesis that the average temperature is $65^{\circ}F$ and collect data a second time while using the same level of significance. In this case, we might be able to claim the null hypothesis is accepted and publish a report.

13.3.3.3 P-value

Another commonly used parameter in hypothesis testing is the **p-value** (or probability value) which is defined as the probability of obtaining test results at least as extreme as the results actually observed during the test, assuming the null hypothesis is correct. More precisely, the p-value is the smallest level of significance that would lead to the rejection of the null hypothesis given the observed data. The p-value does not require imposing a pre-selected level of significance α.

The smaller the p-value, the higher the significance because a smaller p-value indicates to the experimenter that the hypothesis under consideration may not adequately explain the observation.

Figure 55 depicts the p-value (shaded area under the probability distribution). The probability distribution, shown in the figure, is that assumed under the null hypothesis. The shaded area

(which extends indefinitely to the right) represents any value more extreme than the observed data point.

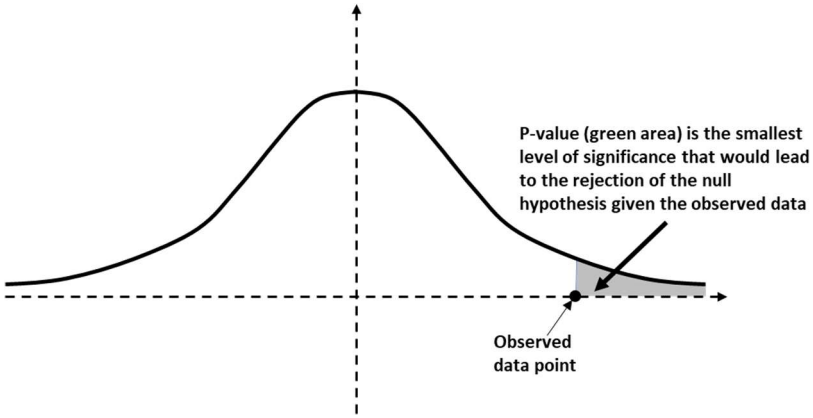


Figure 58. Illustration of p-value

13.3.3.4 P-value versus Level of Significance

Practitioners typically set a conservative standard to meet for the result of an experiment or study to be recognized as significant. The level of significant α is usually set to .05 or a lower value. Under the assumption that the null hypothesis is true, an α set to .05 means the null hypothesis should only be rejected if the observed result is so unusual that it would have occurred by chance at most 5% of the time. The smaller the α, the more stringent the criterion (i.e., the more unlikely it is to find a statistically significant result).

Once the level of significance α has been set, a test statistic is computed. Each test statistic has an associated probability value called a p-value, i.e., the probability of obtaining test results at least as extreme as the results actually observed, assuming that the null hypothesis is correct.

So, the level of significance α is set before the experiment and provides the condition of how extreme the result of an experiment or study must be before the null hypothesis can be rejected, whereas the p-value indicates how extreme the result (as represented by the test statistic) actually is. The p-value is compared to α to determine whether the observed result (test statistic) is significantly different from the null hypothesis:

- If the p-value is less than or equal to α, then the null hypothesis is rejected, and the result is said to be “statistically significant.”
- If the p-value is greater than alpha, then the null hypothesis is not rejected, and the result is said to be “not statistically significant.”

Keep in mind that the null hypothesis is the status quo which we would like to disprove in favor of the alternative. For example, if a pharmacological company has a new drug for diabetes, they would (for example) make a null hypothesis that the new drug offers no advantage over some existing drug and then try to prove via an experiment on trial subjects that the improvements to the trial subjects using the new drug versus those using the existing drug are so pronounced that there is no way the null hypothesis can be true. In general, the goal is to reject the null hypothesis.

13.3.3.5 Example – Large Vat of Ping-Pong Balls

Consider a situation where there is a large vat filled with ping-pong balls of two colors, i.e., red and blue. Your task is to determine if there is roughly an equal number of red and blue balls. The vat is very large and it is not practical to examine each ping-pong ball and count the number of reds and blues. The vat is initially filled by two people, one of whom loads blues and the other loads the reds. The two take turns filling a bucket and dumping the contents into the vat. Each bucket is weighed and brought to an agreed weight by adding or subtracting ping-pong balls. So, in theory, the vat should be about 50% of each color ping-pong ball. However, there is reason to believe that those filling the vat were not following procedures correctly and so, a test is required to verify if the vat does have about 50% of each color ball.

Let the null hypothesis H_0 be that the percentage of red balls is $\mu = 50$ (implying the percentage of blue balls is also 50). The alternative hypothesis H_1 is that the percentage of red balls is not equal to 50.

To test the hypotheses, 20 samples of 100 balls each are drawn from the vat (with replacement after each draw). The results of the samplings are shown in Table 43. To be clear, each entry in the table is the number of blue balls in a sample of 100.

Table 43. Number of Blues in Samples of Size 100

47	44	44	47	49	47	45	48	48	46	49	50	44	44	48	49	49	49	46	45

The mean of the numbers in Table 43 is $\tilde{\mu} = 46.9$. We don't know the standard deviation for the distribution of blue balls in the vat but we can use the sample standard deviation of the numbers in Table 43. The Bessel correction is used (i.e., divide by $n - 1 = 19$ rather than n) in the calculation to get $\hat{\sigma} = 2.023546$.

Assuming a level of significance $\alpha = .05$, we want to determine if the value of the sample mean refutes the null hypothesis or not. For small samples ($n < 30$), it is common to use a test statistic known as the t-test. Associated with the t-test is a distribution defined by the following formula:

$$t = \frac{\tilde{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}$$

where $\tilde{\mu}$ is the sample mean, $\hat{\sigma}$ is an estimate of the standard deviation of the population (i.e., total data space from which samples are drawn), μ is the population mean, and n is the number of data points. Also, as noted earlier, the degrees of freedom for the t-test is $n - 1$.

For the problem at hand, we have

$$t = \frac{\tilde{\mu} - \mu}{\hat{\sigma}/\sqrt{n}} = \frac{46.9 - 50}{2.023546/\sqrt{20}} \cong -6.851153$$

This is basically a standardized value for the mean based on the observed data.

The next step is to use a t-test table (e.g., Table 41 on Page 182) or an online calculator to determine whether the null hypothesis should be rejected or not. For the example under consideration, we have the following results:

Critical t-value range (two-tailed):	[- 2.09302406, 2.09302406]
Two-tailed probability $P(H_0: \mu = 50)$:	0.00000155
Two-tailed probability $P(H_1: \mu \neq 50)$:	0.99999845
p-value:	0.00000155

The critical t-value range is the interval where the null hypothesis should not be rejected. It is basically a confidence interval around a mean of 0 that includes $1 - \alpha$ of the area under the t-distribution with the given degrees of freedom. For the problem at hand, $1 - \alpha = .95$ and the degrees of freedom is 19. Further, our t-value is -6.851153, which is well outside of the critical t-value range and so the null hypothesis should be rejected.

Using the p-value approach, we have that $p = 0.00000155 \leq .05 = \alpha$ and so, the null hypothesis is rejected. This is highly suggestive that the alternative hypothesis is true.

Let's reevaluate our hypothesis and test against a revised population mean $\mu = 47$, i.e., assume the actual number of blue balls (out of 100) is 47. In this case, we get the following results:

t-statistic:	-0.22100556
Critical t-value range (two-tailed):	[- 2.09302406, 2.09302406]
Two-tailed probability $P(H_0: \mu = 47)$:	0.82744578
Two-tailed probability $P(H_1: \mu \neq 47)$:	0.17255422
p-value:	0. 82744578

This time, our test statistic does fall within the critical t-value range. So, we do not reject the null hypothesis in this case.

Using the p-value approach, we have that $p = 0.82744578 \geq .05 = \alpha$ and so, the null hypothesis is not rejected.

13.3.3.6 Example – Comparing Battery Type Lifetimes

Consider the problem of comparing the average lifetimes of two types of batteries (Battery Type X and Battery Type Y). Thirty data points for instances of Battery Type X and Battery Type Y are listed below. Each data point represents the lifetime (in months) of a battery instance. The maker of Battery Type X would like to claim their battery lasts on average 3 or more months longer than batteries of Type Y.

Battery Type X

29.9389	28.6328	31.5358	29.5931	25.2887	28.1834
25.6264	30.8459	20.3379	27.4113	23.1402	27.8455
19.7259	30.2547	35.4607	26.3644	26.8253	34.5251
29.3185	29.9598	26.2742	31.0308	27.1553	25.9372
28.5168	26.6379	31.6680	25.8408	27.5514	27.9000

Battery Type Y

23.6183	22.9001	26.7005	21.5633	22.8037	22.4295
29.2587	24.0003	21.0254	25.4782	22.5474	23.0955
16.1366	22.3171	22.4769	17.1468	22.1785	20.2322
20.8528	16.6673	22.9132	28.8552	20.4779	21.0232
26.0684	24.0126	22.5906	25.0016	26.5045	22.3598

We start by stating the null and alternative hypotheses:

$H_0: \mu_X \leq \mu_Y + 3$ where μ_X and μ_Y are the means of Battery Types X and Y, respectively

$H_1: \mu_X > \mu_Y + 3$

Something called the two-sample t-test [67] can be used here. Microsoft Excel has a data analysis package which performs the two-sample t-test. The above data was entered into Excel, with the following results:

Table 44. Results of two-sample t-test for battery lifetimes

	Battery Type X	*Battery Type Y*
Mean	27.97755974	22.7745311
Variance	11.79301461	9.39436515
Observations	30	30
Pooled Variance	10.59368988	
Hypothesized Mean Difference	3	
df	58	
t Stat	**2.621453425**	
P(T<=t) one-tail	**0.005581338**	
t Critical one-tail	**1.671552762**	
P(T<=t) two-tail	0.011162676	
t Critical two-tail	2.001717484	

For the problem at hand, the main items of interest are highlighted in bold in Table 44.

Since the "t Stat" value of 2.6215 is greater than (i.e., more extreme than) the "t Critical one-tail" value of 1.6716, the null hypothesis is rejected.

Using the p-value approach, the p-value is 0.00558 which is less than $\alpha = .05$, it is again concluded that the null hypothesis should be rejected.

The conclusion is that the results of the experiment are strongly suggestive of the alternative hypothesis being true.

13.3.3.7 Example – Type II Errors

Consider a new material being planned for manufacturing. The manufacturer doesn't want to produce and sell the new material unless it has a strength rating of at least 25 (on a scale from 1 to 50). The manufacturer hires a testing company which, in turn, tests 50 samples. The tests resulted in a sample mean of 25.9 and a sample standard deviation of 4.3.

The following null and alternative hypotheses are posed:

$$H_0: \mu \leq 25$$

$$H_1: \mu > 25$$

Select the level of significant to be .05. We use the z-score calculator from https://www.calculator.net/z-score-calculator.html, noting that what they call "raw score, x" is $\tilde{\mu}$ in the context of the example.

The following is entered into the z-calculator:

Raw Score x, i.e., $\tilde{\mu}$	**25.9**
Population Mean μ:	**6**
Standard Deviation $\tilde{\sigma}/\sqrt{50}$:	$\frac{4.3}{\sqrt{50}} = \mathbf{.608}$

The output from the calculator is as follows:

Z-score = 1.48026		
Probability of $\tilde{\mu} < 25.9$:	0.9306	
Probability of $\tilde{\mu} > 25.9$:	0.069402	(this is essentially the p-value)
Probability of $25 < \tilde{\mu} < 25.9$:	0.4306	

We see from the above output that the p-value is .069402 which is greater than $\alpha = .05$ and so, we do not reject the null hypothesis and thus, do not conclude that the new material has an average strength rating greater than 25.

Regarding the Type II error, it is critical to note that the alternative hypothesis covers a range of possible values, and for each possible value of the alternative hypothesis, there will be a different value for the probability of a Type II error. If we choose (as an alternative hypothesis) the mean to be 26 (which is the critical value (c.v.) of the sample mean corresponding to $\alpha = .05$), then we get the situation shown in Figure 59. The normal curve on the left is the distribution corresponding to the null hypothesis and the normal curve on the right is the distribution corresponding to the alternative hypothesis under the assumption that 26 is the actual mean. The dark gray area is the rejection region for the null hypothesis. If the null hypothesis is correct and the sample mean falls in the light gray region, then the null hypothesis will not be rejected. This is a Type II error. The area of the light gray region is the probability of a Type II error, which in this case is $\beta = .5$.

If we choose a larger value for the mean of the alternative hypothesis (26.41) and assume that it is in fact true, then we get the situation shown in Figure 60. Now the probability of a Type II error is $\beta = .25$ (area of the light gray region). In general, as the chosen value for the mean of the distribution corresponding to the alternative hypothesis moves further away from the assumed mean of the distribution corresponding to the null hypothesis distribution, the probability of a Type II error becomes less and eventually approaches 0.

If we choose a smaller value for the mean of the alternative hypothesis (25.59) and assume that it is in fact true, then we get the situation depicted in Figure 61. The probability of a Type II error is $\beta = .75$ (area of the light gray region). In general, as the chosen value for the mean of the distribution corresponding to the alternative hypothesis moves close to the assumed mean of the distribution corresponding to the null hypothesis distribution, the probability of a Type II error becomes greater and eventually approaches 1.

If we chose the mean of the alternative hypothesis to be the same as the sample mean, i.e., 25.9, then $\beta = .5654$.

(Credits: The following three graphs were generated using an online application at https://shiny.rit.albany.edu/stat/betaprob/).

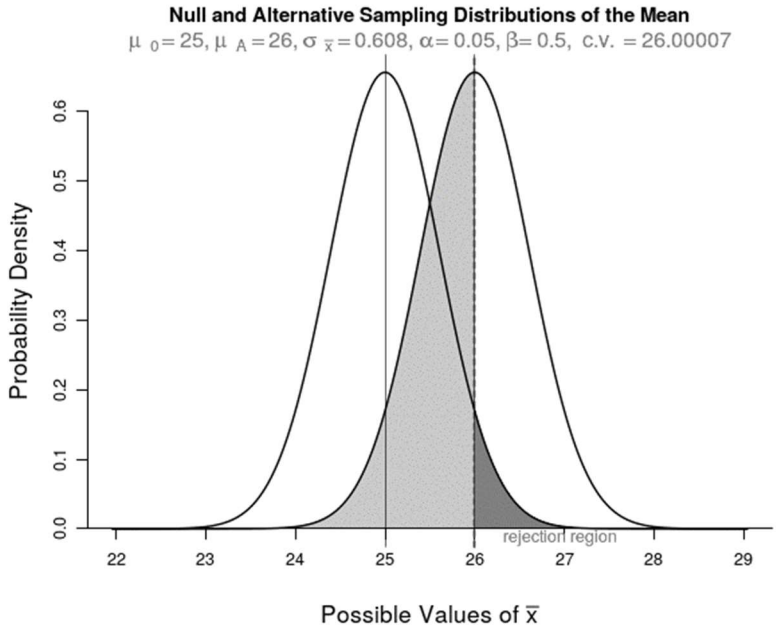


Figure 59. Type II Error of .5

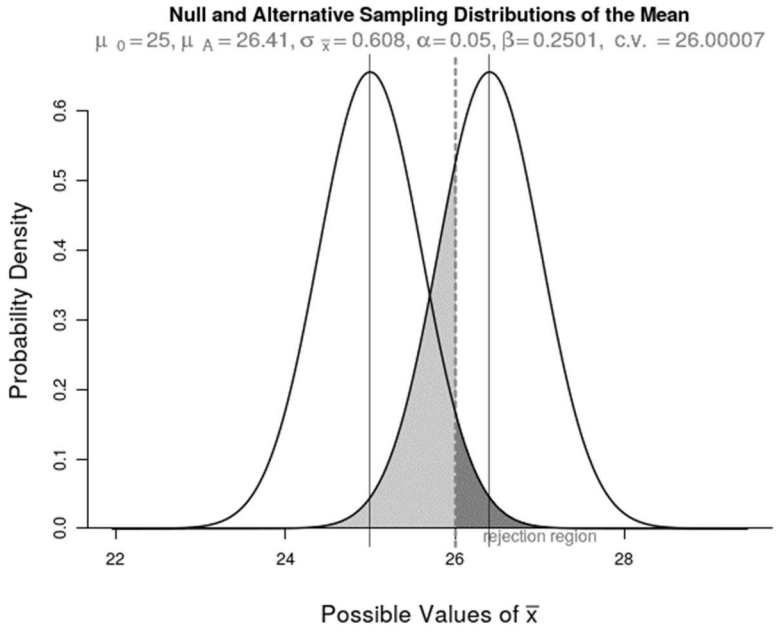


Figure 60. Type II Error of .25

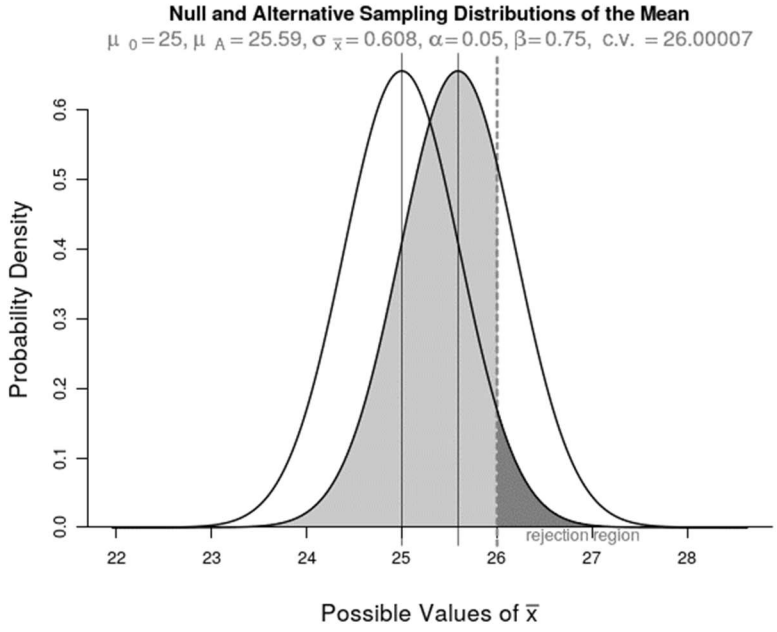


Figure 61. Type II Error of .75

13.3.4 Relationship between Confidence Intervals and Hypothesis Testing

Confidence intervals and two-sided hypothesis tests lead to the same conclusions. A two-sided hypothesis test is equivalent to the act of computing a confidence interval for the sample parameter (e.g., a mean) and determining whether the confidence interval contains the stipulated (hypothesized) value for the sample parameter.

- The null hypothesis is not rejected if the stipulated value for the given parameter falls within the confidence interval for the sample parameter.
- The null hypothesis is rejected if the stipulated value for the given parameter falls outside of the confidence interval for the sample parameter.

However, this equivalence to confidence intervals does not hold for one-sided hypothesis tests.

The YouTube video entitled "Hypothesis Test vs. Confidence Interval | Statistics Tutorial #15 | MarinStatsLectures" [68] provides a good visual presentation of the relationship between confidence intervals and two-sided hypothesis tests.

13.3.5 Regression Analysis

Regression analysis entails usage of a set of statistical tools to estimate the relationships between a dependent variable (also known as the "outcome" or "outcome variable") and one or more independent variables (also known as "predictors"). The relationships can be linear where the dependent variable (y) is represented as a linear combination of the independent variables $(x_1, x_2, \dots, x_n)$ given by the formula $y = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ such that $a_i, i = 1,2, \dots, n$ are constants. It is also possible to have a non-linear relationships among the independent and dependent variables, e.g., $y = 7{x_1}^3 - 3{x_2}^{-1} + 17e^{x_3} - 11\ln(x_4)$.

When employing regression analysis, data points are used to determine an equation that relates the dependent variable to the independent variables. The equation is then used to predict values for the dependent variable based on various values of the independent variables.

13.3.5.1 Simple Linear Regression

The simplest version of regression analysis entails the case where there is one dependent variable and one independent variable, and the two are related by a straight line. The line is constructed so as to minimize the sum of the squares of the vertical distances from each data point to the line (this is called a **least squares** fitting of a line to a set of data points). Figure 62 depicts a dashed line fitted to a set of data points using the least squares concept (distances shown as short vertical lines in the figure).

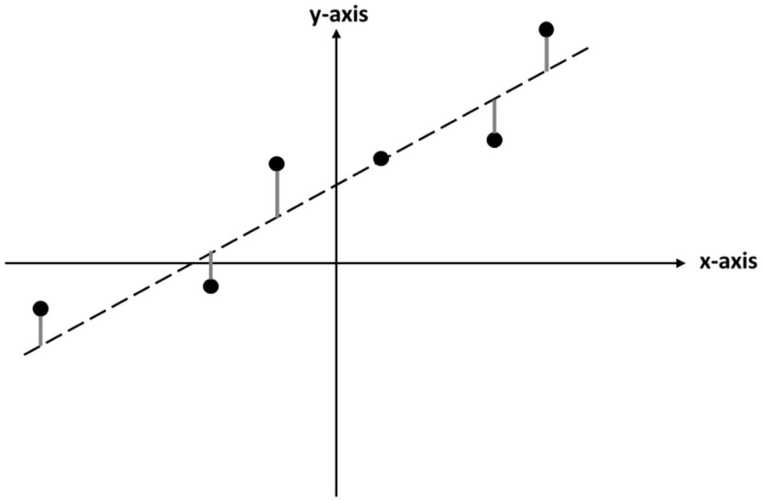


Figure 62. Least Squares Fitting – Conceptual Drawing

For a set of data points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, it is possible prove (see [69]) that the least squares line is given by $y = ax + b$ where

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

As an example, take the data points concerning the high school and college Grade Point Averages (GPAs) for a set of students (show in Table 45). The dependent variable (y) is the college GPA and the independent variable (x) is the high school score. Each (x, y) pair represents one student.

Table 45. Comparison of High School and College GPAs

High School (x)	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3	3.1	3.3	3.5	3.7	3.9	4
College (y)	2.8	3	2.2	2.5	3	3	2.7	3.5	2.8	3.1	3.3	4	3.6	3.6

Figure 63 shows the least squares line (and associated equation) that fits the given data (generated using Microsoft Excel).

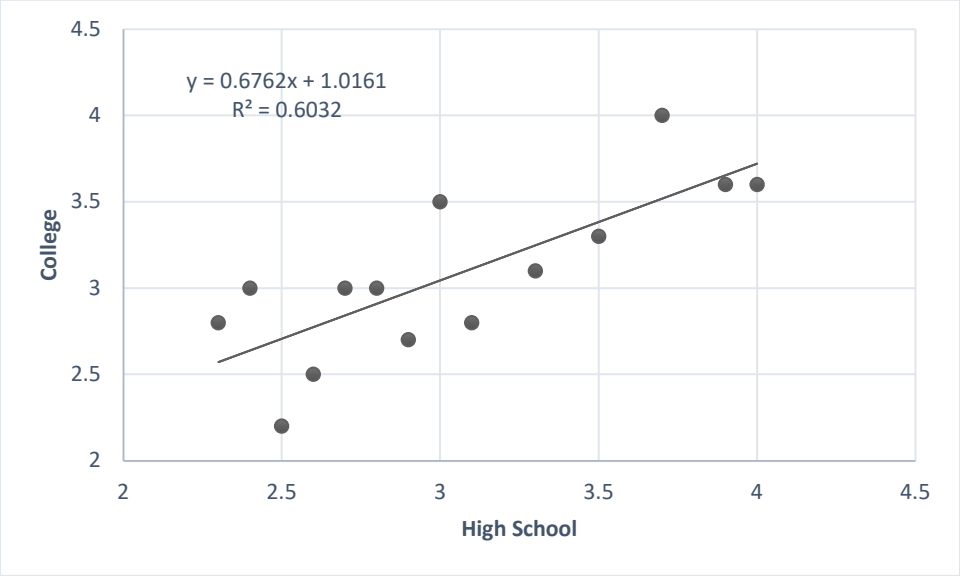


Figure 63. Least Squares Line for GPA Comparison

The most common measure of how well a set of data points fits with the associated least squares line is called the **correlation coefficient**. The value of the correlation coefficient is between -1 and 1, with the following interpretations:

- The closer the correlation coefficient is to 1, the stronger the positive relationship, and the close the correlation coefficient is to -1, the stronger the negative relationship.
- Correlation coefficient values less between -0.8 and 0.8 are not considered significant.

The formula for the correlation coefficient is as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

For the GPA example, we are given the square of the correlation coefficient (i.e., R^2 in Figure 63). So, taking the square root, we get .777 as the correlation coefficient.

It is critical to understand that a high correlation does not imply causality. For example, there could be a third variable z that strongly impacts x and y. The sale of bathing suits (x) and the sale of packaged ice (y) may be strongly correlated but one does not cause the other. It is a third variable (hot summer weather) that impacts (causes) both x and y.

13.3.5.2 Multiple Linear Regression

In multiple linear regression (or just multiple regression) several independent variables are used to predict the outcome of a dependent variable, where the dependent variable is represented as a linear combination of the independent variables.

As an example, consider the setting of a price (y) for a used car based on mileage (x_1), age (x_2) and make of car (x_3), with the goal to set the price so that the car will sell quickly. Assume the make of the car can be converted to a number from 1 to 10 (with 1 having the lowest resale value and 10 having the highest). The assumption is that y can be represented as a linear combination of x_1, x_2 and x_3, i.e., $y = a_1x_1 + a_2x_2 + a_3x_3 + b$ where a_1, a_2, a_3 and b are constants to be determined. The constant b could be, for example, the scrap value of the car.

13.3.5.3 Nonlinear Regression

In nonlinear regression several independent variables are used to predict the outcome of a dependent variable, where the dependent variable is represented as a nonlinear combination of the independent variables.

Consider the following example where we hypothesize that the relationship between an independent variable x and a dependent variable y is given by $y = x^2 + 1$. In order to confirm our theory, several data samples are collected as shown in Table 46.

Table 46. Nonlinear Regression Example

Independent Variable (x)	1	2	3	4	5	6
Dependent Variable (y)	2.1	4.8	10.15	17.3	25.95	35.9

If we fit a second degree polynomial to the data points (using the Trending capability from Microsoft Excel), we get the result shown in Figure 64.

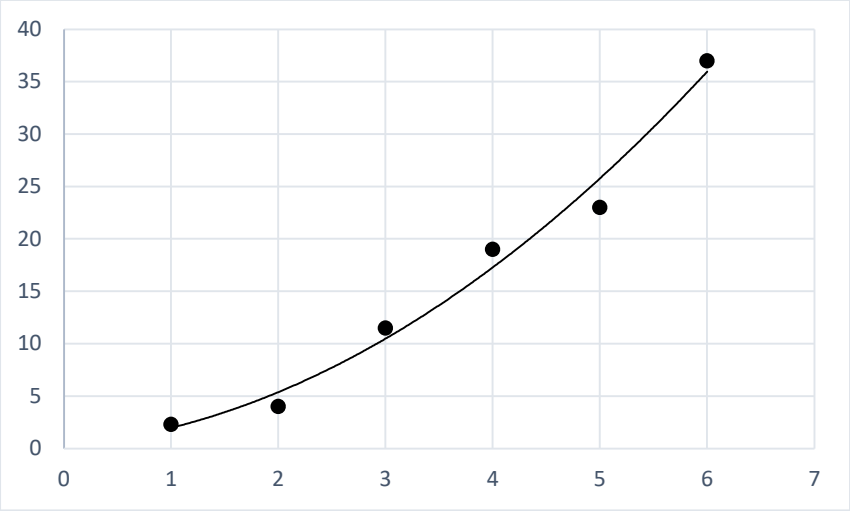


Figure 64. Nonlinear Regression using a Second Degree Polynomial

If we use a 6th degree polynomial, we can get an exact fit of the curve to the data points, as shown in Figure 65. However, this may be an example of **overfitting**. Overfitting occurs when a model too closely or exactly fits a particular set of data, and consequently, may not be a good fit for other data nor predict future observations accurately.

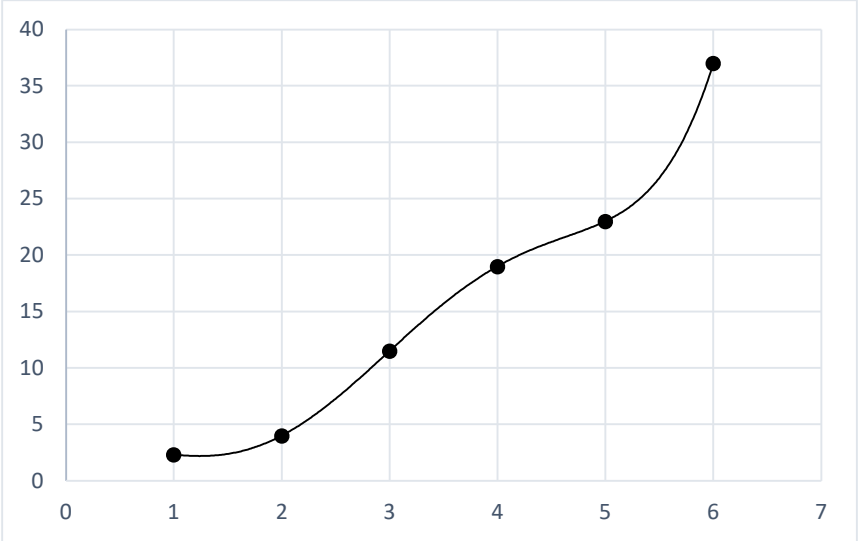


Figure 65. Nonlinear Regression using a Sixth Degree Polynomial

13.3.5.4 Autoregression

Autoregression is a type of regression where the dependent variable depends on one or more past values of itself. In other words, a model is considered to be autoregressive if it predicts future values based on past values.

For example, an autoregressive model might be used to predict a stock's future prices based on its past prices. Assume the closing prices for the last three days are used to predict a stock's price on the following day. Let y be the dependent variable (today's, yet unknown, price for a given stock)

and let x_1, x_2 and x_3 be the stock price 1, 2 and 3 days ago, respectively. The implied relationship among the dependent and independent variables can be represented as follows:

$$y = a_1x_1 + a_2\,x_2 + a_3x_3 + b \text{ where } a_1, a_2, a_3 \text{ and } b \text{ are constants.}$$

The prices for the given stock on 13 consecutive trading days are shown in the following table:

1	2	3	4	5	6	7	8	9	10	11	12	13
12.25	12.10	11.95	12.50	12.35	12.15	12.05	12.00	11.95	11.80	12.10	12.35	12.70

Starting from Day #4, we have stock prices for the past 3 days, as shown in the following table:

Day	Current (y)	1 Day Ago (x_1)	2 Days Ago (x_2)	2 Days Ago (x_3)
4	12.50	11.95	12.10	12.25
5	12.35	12.50	11.95	12.10
6	12.15	12.35	12.50	11.95
7	12.05	12.15	12.35	12.50
8	12.00	12.05	12.15	12.35
9	11.95	12.00	12.05	12.15
10	11.80	11.95	12.00	12.05
11	12.10	11.80	11.95	12.00
12	12.35	12.10	11.80	11.95
13	12.70	12.35	12.10	11.80

Using the regression capability from the Microsoft Excel Data Analysis tool set, we get the following equation relating the dependent variable to the three independent variables:

$$y = .568x_1 - .259x_2 - .322x_3 + 12.348$$

The Excel regression tool also provides the **coefficient of determination**, denoted R^2, which is the proportion of the variance in the dependent variable that is predictable from the independent variables. A value of R^2 close to 1 indicates a good fit of the model to the data. For the problem at hand, $R^2 = .331$ and so, using autoregression on the previous three days of this particular stock price appears not to be a good approach.

13.4 Statistical Paradoxes

The following paradoxes have arisen in actual statistical experiments. They are stated here for the reader's entertainment as well as a warning that one needs to be very careful with the interpretation of statistics.

13.4.1 Berkson's Paradox

13.4.1.1 Conditional Dependence

Conditional (or negative) dependence is a relationship between two or more events that become dependent when a third event occurs. In Figure 66, events A and B are independent but the occurrence of A or B increases the likelihood that C will occur. Suppose that C occurs. Further, assume the occurrence of event B decreases the probability of the occurrence of event A.

(Similarly, event A occurring will decrease the probability of the occurrence of B). In this situation, the two events A and B are said to be conditionally (or negatively) dependent on each other because the probability of occurrence of each is negatively dependent on whether the other occurs. In terms of probabilities, we have

$$P(A|C,B) < P(A|C)$$
$$P(B|C,A) < P(B|C)$$

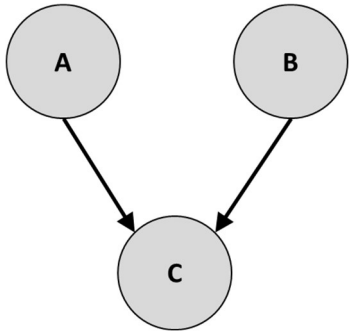


Figure 66. Conditional Dependency

The concept of conditional dependency is at the heart of Berkson's paradox. Specific examples of Berkson's paradox are described in the following sections.

13.4.1.2 Restaurant Example

A food critic claims that the restaurants she visits, in a given town, have either an uninteresting menu and adequate parking space, or vice versa. Our food critic thinks it is strange that there is an inverse correlation between an uninteresting menu and adequate parking, and vice versa. In other words, when she visits a restaurant with an interesting menu, the parking is usually inadequate, and when she visits a restaurant with adequate parking, the menu is usually uninteresting. The confusion arises from the fact that the food critic is unlikely to visit restaurants that have both an uninteresting menu and inadequate parking.

Let A be the event that the food critic picks a restaurant with an interesting menu and B be the event that the food critic picks a restaurant with adequate parking.

Assume that events A and B are independent, i.e., $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Further, A and B are conditionally dependent with respect to the event $A \cup B$, i.e., $P(A|B, A \cup B) < P(A|A \cup B)$. Relative to the definition of conditional dependency in the previous section, $C = A \cup B$.

$P(A|B, A \cup B)$ is defined to be $P(A|B \cap (A \cup B))$. Also, note that $B \cap (A \cup B) = B$ (absorption law from Theorem 7-5) and so

$$P(A|B, A \cup B) = P\big(A\big|B \cap (A \cup B)\big) = P(A|B) = P(A) \qquad \text{(Equation 1)}$$

For the sake of argument, let $P(A) = .5$ and $P(B) = .5$. We have that

$$P(A \cap B) = P(-A \cap B) = P(A \cap -B) = P(-A \cap -B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

$$P(A|A \cup B) = \frac{P\big(A \cap (A \cup B)\big)}{P(A \cup B)} = \frac{P(A)}{P(A \cup B)} = \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}$$

So, $P(A|A \cup B) = \frac{2}{3} > \frac{1}{2} = P(A) = P(A|B) = P(A|B, A \cup B)$, using Equation 1. Similarly, $P(B|A \cup B) = \frac{2}{3} > \frac{1}{2} = P(B) = P(B|A) = P(B|A, A \cup B)$. Thus, A and B are conditionally dependent with respect to $A \cup B$.

In words, $P(A|A \cup B)$ is the probability that the food critic selects a restaurant with an interesting menu given that a restaurant with an interesting menu or with adequate parking is selected. **This excludes the case of $-A$ (selecting a restaurant with uninteresting menu) and $-B$ (selecting a restaurant with inadequate parking)!** On the other hand, $P(A) = .5$ is the unconditional probability of selecting a restaurant with an interesting menu which includes the possibility that the restaurant also has adequate parking.

Table 47 provides a visual representation of the restaurant example. In the table, an equal number of restaurants of each combination is assumed and so, $P(A) = P(B) = \frac{1}{2}$. Now, $P(A|A \cup B)$ is the probability of A when restricted to the gray area in the table which is the sum of the entries in the left column divided by the sum of the entries in the gray area, i.e., $\frac{10+1}{10+10+10} = \frac{2}{3}$. However, we get a lower probability if we compute $P(A|B, A \cup B) = P(A) = \frac{10+10}{10+10+10+10} = \frac{1}{2}$. Thus, knowing that B has occurred lowers the probability of A, given $A \cup B$. This is because $P(A|A \cup B)$ excludes the case of $-A \cap -B$.

Table 47. Restaurant Visit Example

	interesting menu *A*	**uninteresting menu -*A***
adequate parking *B*	10	10
inadequate parking -*B*	10	10

A paradox arises because the conditional probability of A given B within the three-cell subset equals the conditional probability in the overall population, i.e., $P(A|B, A \cup B) = P(A|B) = P(A) = \frac{1}{2}$, but the unconditional probability within the subset, $P(A|A \cup B) = \frac{2}{3}$, is inflated relative to the unconditional probability in the overall population, hence, within the subset, the presence of B decreases the conditional probability of A (back to its overall unconditional probability).

The number of restaurant types does not need to be the same for all categories for the paradox to arise, as is illustrated in Table 48. Events A and B are still independent since $P(A) = P(A|B) = \frac{1}{13} \cong .077$ and $P(B) = P(B|A) = \frac{1}{8} = .125$. However, the unconditional probabilities within the subset $A \cup B$ are inflated as before, i.e., $P(A|A \cup B) = \frac{80}{200} = .4$ and $P(B|A \cup B) = \frac{130}{200} = .65$.

Table 48. Restaurant Visit Example Modified

	interesting menu A	uninteresting menu -A
adequate parking B	10	120
inadequate parking -B	70	840

13.4.1.3 Stamp Example

The Wikipedia article on this topic [70] provides another interesting example concerning pretty and rare stamps. From the Wikipedia article:

> As a quantitative example, suppose a collector has 1000 postage stamps, of which 300 are pretty and 100 are rare, with 30 being both pretty and rare. 10% of all his stamps are rare and 10% of his pretty stamps are rare, so prettiness tells nothing about rarity. He puts the 370 stamps which are pretty or rare on display. Just over 27% of the stamps on display are rare (100/370), but still only 10% of the pretty stamps are rare (and **100% of the 70 not-pretty stamps on display are rare**). If an observer only considers stamps on display, they will observe a spurious negative relationship between prettiness and rarity as a result of the selection bias (that is, not-prettiness strongly indicates rarity in the display, but not in the total collection).

Table 49 shows the breakdown of the various stamp categories based on the description from the Wikipedia article. The gray cells represent the stamps put on display.

Table 49. Breakdown of Stamps per Category

	rare	not rare	
pretty	30	270	300
not pretty	70	630	700
	100	900	1000

13.4.2 False Positive Paradox

Consider an infectious disease test on a population of 10000 people where 1% of the people are actually infected. The test has a false positive rate of 5% and false negative rate of 5%.

Given the infection rate of 1%, the expectation is that 100 people are infected and 9900 are not infected. Of the 100 infected people, the expectation is that 5% will receive a false negative result (i.e., 5 people). Of the uninfected people, the expectation is that 5% will receive a false positive result (i.e., 495 people). The expected results are summarized in Table 50.

Table 50. False Positive Paradox - Example

	A: Infected	**-A: Uninfected**	**Total**
B: Tested positive	95 (true positive)	*495* (false positive)	590
-B: Tested negative	5 (false negative)	9405 (true negative)	9410
Total	100	9900	10000

The probability that someone is told they are infected (as a result of the test) and in fact are infected is $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{95}{590} \cong .16$ and herein lies the paradox, i.e., even though the test has low rates for the two types of errors, there is only about a 16% probability that someone is actually infected given that they tested positive. The confusion arises from the fact that only 1% or about 100 people out of 10000 are infected and the test (while relatively accurate) is applied to 9900 people who are uninfected (resulting in an expected number of 495 false positives which is much larger than the total expected number of infected people, i.e., only 100).

In general, the accuracy of a test needs to be much better than the prevalence rate of the trait under consideration to give a reasonable distinction between those with and without the given trait.

13.4.3 Simpson's Paradox

Simpson's paradox is a phenomenon in which a trend appears in several different groups of data but the trend reverses when the groups are combined. For example, consider two different pain relief medications A and B, and their effect on headaches and muscle aches. As shown in Table 51, Medication A is more effective in relieving headaches than Medication B, and Medication A is also more effective in relieving muscle aches than Medication B. However, if we combine the results, Medication B is overall more effective in relieving pain than Medication A. (The fractions in the table, e.g., $95/100$, should be read as "the medication was given to 100 patients and 95 indicated complete relief or significant reduction in pain.") The problem comes from the fact that the medication is not given to equal numbers of people for the various cases.

Table 51. Effects of Pain Relief Medication

	Medication A	Medication B
Headaches	$\frac{95}{100} = \mathbf{95}\%$	$\frac{750}{1000} = 75\%$
Muscle aches	$\frac{300}{500} = \mathbf{60}\%$	$\frac{25}{50} = 50\%$
Total	$\frac{395}{600} = 66\%$	$\frac{775}{1050} = \mathbf{74}\%$

Simpson's paradox is an example of a veridical paradox.

13.5 Exercises

1. Let random variable X represent the maximum amount of weight that female high school seniors can deadlift. Assume 250 students are randomly selected and their maximum deadlift is recorded. The standard deviation is known to be $\sigma = 11$ kgs. The sample mean $\tilde{\mu} = 65$ kgs. Use the normal distribution (z-value) to find a 98% confidence interval for the actual mean μ. ***Answer**: (63.382, 66.618).*

2. Use the Student's t-distribution to find a 95% confidence interval for the mean of a random variable X, given the following random samples from X:

1.248	1.206	1.227	1.226	1.294	1.225	1.292	1.294	1.275	1.203
1.207	1.243	1.221	1.246	1.276	1.25	1.217	1.234	1.276	1.289

   ***Hint**: Use the formula in Section 13.3.2.2 since we have less than 30 sample points.* ***Answer**: (1.233, 1.262).*

3. Based on 1000 samples (with replacement) from a container with several million white and black marbles (with marbles of no other colors), the sample proportion of white marbles was .55. Find a 95% confidence interval for the proportion of white marbles. ***Hint**: Use the formula from Section 13.3.2.3.* ***Answer**: (.519, .581).*

4. In the previous exercise, how large does the sample size need to be to get a 95% confidence interval of $\pm.02$ about the mean. ***Answer**: 2377.*

5. Find the least squares line and associated correlation coefficient for the following set of data points:

1	2	3	4	5	6	7
4.1	2.9	2	0.9	-0.05	-1.05	-1.95

6. A car manufacturer claims that their updated sports car can accelerate from 0 to 60 miles per hour (mph) in 2.5 seconds. An independent testing company has compiled 30 test runs of the car with the results listed in the table below. Do an test to determine the validity of the hypothesis that the mean acceleration time from 0 to 60 mph is 2.5 seconds with $\alpha = .05$ level of significance. ***Hint**: Since we are on the borderline (n=30) of whether to use the z-test or t-test, try using both.*

2.523	2.505	2.453	2.503	2.493	2.553	2.563	2.535	2.438	2.533
2.548	2.500	2.465	2.538	2.470	2.533	2.493	2.520	2.493	2.558
2.513	2.475	2.513	2.555	2.480	2.558	2.555	2.453	2.450	2.473

7. For one interval lag (i.e., autoregression using the previous data point), determine the coefficient of determination R^2 for the following (ordered) data sample:

4.000	5.250	4.889	5.063	4.960	5.028	4.980	5.016	4.988	5.010
4.992	5.007	4.994	5.005	4.996	5.004	4.997	5.003	4.997	

14 Graph Theory

14.1 Basic Concepts

Graph theory is a branch of mathematics concerned with configurations of connected points. The beginnings of graph theory can be traced to recreational math problems (see the Königsberg bridge problem in Section 14.5), but it has grown into a prominent area of mathematical research. For example, graph theory is being actively used in fields such as biochemistry (genomics), electrical engineering (communication networks and coding theory), computer science (algorithms and computation) and operations research (scheduling).

A graph is an abstraction that represents the interrelationships among a system of things. More formally, a **graph** G is a finite nonempty set V of objects called **vertices** along with a set E consisting of 2-element subsets of V each of which is called an **edge**.

For example, let G be a graph consisting of airports A, B, C and D (the vertices) and connections (i.e., available flights) among the following pairs of airports (the edges):

$$\{(A,B)_1, (A,B)_2, (A,C), (A,D), (B,C), (B,D), (C,D), (D,D)\}.$$

The flight from D to D is an aerial sightseeing tour that leaves from D and returns back to D. There are two different flights between A and B, labeled as $(A,B)_1$ and $(A,B)_2$. The graph G can be represented in the diagram shown in Figure 67. The intersection between (A,C) and (B,D) should not be confused as a vertex – it is not.

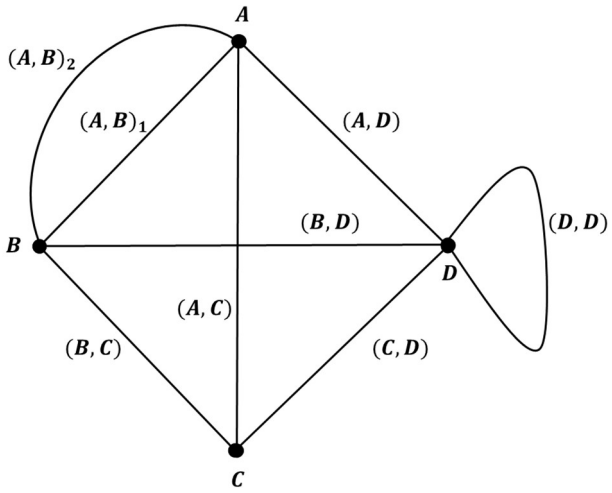


Figure 67. Airport – Flights Graphs

The set V, sometimes written $V(G)$, is the vertex set of G. The set E, sometimes written as $E(G)$, is the edge set of G. G can be written as $G = (V, E)$ to highlight the associated vertex and edge sets.

The **order** of G is the number of vertices in $V(G)$. The **size** of G is the number of edges in $E(G)$. For the example in Figure 67, the order of G is 4 and the size of G is 8.

Vertices u and v are said to be **adjacent** if there is an edge between them. The **degree of a vertex** v, denoted $deg(v)$, is the number vertices adjacent to v. In Figure 67, $deg(A) = \deg(B) = 4$, and $deg(C) = 3$. A loop adds two to the degree of a vertex. So, $deg(D) = 5$.

If the discussion is restricted to **simple graphs**, i.e., graphs that do not have loops (such as D to D in the previous example) and which have at most one edge between any pair of vertices, then the following theorems hold true. Graphs that do have loops or multiple edges between pairs of vertices are called **multigraphs**.

Theorem 14-1 If G is a simple graph of order n and size m, and with vertices $v_1, v_2, \ldots, v_n$, then $\sum_{i=1}^{n} deg\ (v_i) = 2m$.

Proof: When the degrees of the vertices of any simple graph G are added, each edge of G is counted twice■

Theorem 14-2 Every simple graph G has an even number of vertices of odd degree. (Note that "even number" includes the possibility of zero.)

Proof: Keeping in mind Theorem 14-1, we know that the sum of the degree must add to an even number. Divide the vertices of G into two sets A (vertices with even degree) and B (vertices with odd degree). The sum of the degrees of the vertices in set A is clearly an even number since the sum of even numbers is again even. If the number of vertices in B is odd, then the sum of the degrees of the vertices in B is also odd since the sum of an odd number of odd numbers is an odd number. However, this leads to the sum of the degrees of the vertices in A and B being odd, which contradicts Theorem 14-1. Thus, B must have an even number of vertices. Since A and B each have an even number of vertices, so does G■

The complement of a simple graph G is the graph G' which has the same vertices as G but none of the same edges. In other words, G' only has edges between vertices that are not adjacent in G, and vice versa. Figure 68 depicts a graph G and its complement G'. Of course, the complement of a complement takes you back to the original graph.

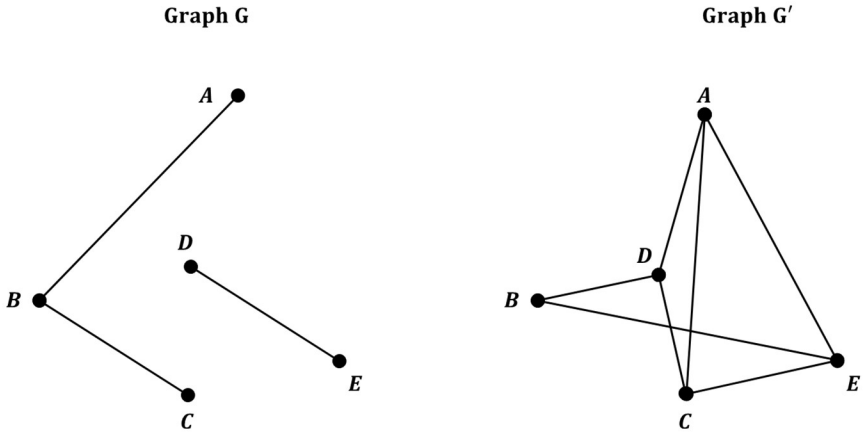


Figure 68. Complementary Graphs

A graph G is **connected** if it is possible to walk (i.e., traverse edges) from any vertex to any other vertex of G. Otherwise, G is **disconnected**. In Figure 68, Graph G is connected and Graph G' is disconnected.

A disconnected graph can be divided into two or more connected graphs, called **components**. In Figure 68, Graph G has two components.

All the graphs in this section are undirected graphs (edges can be traversed in either direction). While not discussed here, it is also possible to have directed graphs where some of the edges can only be traversed in one direction.

14.2 Classification

The discussion in this section is restricted to simple graphs. The qualification of "simple" will be omitted.

14.2.1 Almost Irregular Graphs

An **irregular graph** is one in which each vertex is of a different degree from all other vertices in the graph. As hinted in the title of this section, the set of irregular graphs is empty. This fact is proven in the following theorem.

Theorem 14-3 There are no irregular graphs of order 2 or more.

Proof: Assume that there is an irregular graph of order $n \geq 2$. This means that each vertex is of a different degree. Thus, the only possible degrees for a vertex in a graph of order n are $0, 1, 2, \ldots, n-1$. So, for G to be irregular, there must be exactly one vertex of degree $0, 1, 2, \ldots, n-1$. However, the vertex of degree $n-1$ must be adjacent to all the other vertices, but the vertex of degree 0 is adjacent to no other vertex. Thus, we have arrived at a contradiction and the initial assumption of G being irregular must be false■

The above theorem is related to the old puzzle concerning n people at a party where for each 2 people, they are either friends or not friends. Let the people be vertices in a graph. Further, let there be an edge between two people (vertices) who are friends, and no edge when two people are not friends. By Theorem 14-3, there are at least two people at the party who have the same number of friends (at the party).

There are graphs that are **almost irregular**, i.e., a graph that has exactly one pair of vertices of the same degree. Graphs G and H in Figure 69 are both almost irregular. For Graph G, only vertices A and B have the same degree (2 in this case). For Graph H, only A and B have the same degree (1 in this case).

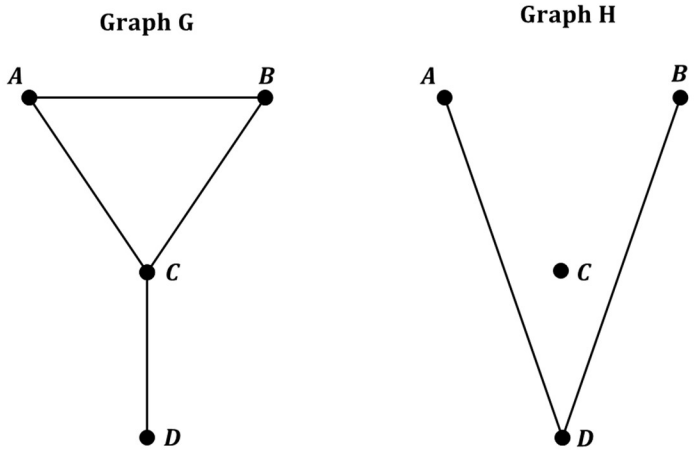


Figure 69. Almost Irregular Graphs

Graph G and H in Figure 69 are complements. It turns out that this is not a coincidence. In fact, we have the following theorem (the proof of which is omitted).

Theorem 14-4 For $n \geq 2$, there are exactly two almost irregular graphs of order n. These graphs are complements of each other.

For every given set of positive integers whose largest member is n, there is a graph of order $n + 1$, the degrees of whose vertices are precisely these integers. The possible graphs for the case of $n = 4$ are shown in Figure 70. The general result is captured in the following theorem (the proof of which is omitted).

Theorem 14-5 For every set of positive integers S whose largest value is n, there exists a graph G of order n + 1 such that degrees of the vertices of G exactly comprise the set S.

Keep in mind that the set S must have n as an element. By the definition of a set, S does not have any repeated elements. However, as can be seen in Figure 70, it is possible to have repeats in terms of the number of vertices that have a given degree.

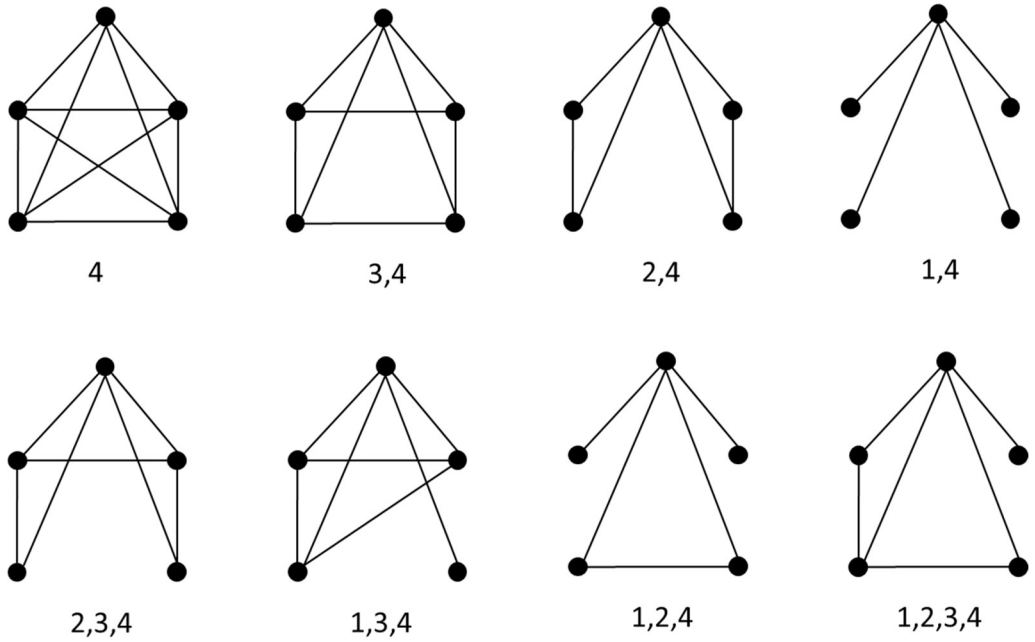


Figure 70. All possible graphs of order 5 with a vertex of degree 4

14.2.2 Regular Graphs

A graph G is classified as **regular** if all its vertices are of the same degree. If the common degree is r, then G is said to be r-regular. If G is an r-regular graph of order n, then the following inequality must hold $0 \leq r \leq n - 1$.

A 0-regular graph of order n is simply a set of n vertices with no edges.

Some types of regular graphs arise frequently and are given specific names, as noted below.

The graph of order n where all possible edges are included is called a **complete graph** of order n and denoted by K_n. K_n is an $(n - 1)$-regular graph. Some examples of complete graphs are shown

in Figure 71. The vertices are only the small solid black circles and not the interior intersection of the edges.

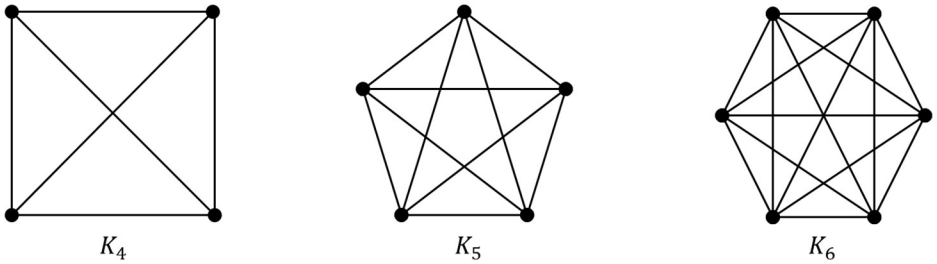


Figure 71. Examples of Complete Graphs

The complete graph of order n where all the vertices are of order two and connected in single cycle is called the **cyclic graph** of order n and denoted by C_n. Figure 72 shows some examples of cyclic graphs.

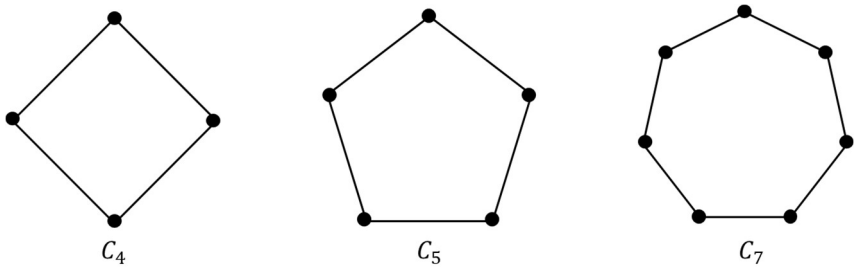


Figure 72. Examples of Cyclic Graphs

By Theorem 14-2, no graph can contain an odd number of odd vertices. Thus, there are no r-regular graph of order n where both r and n are odd numbers. Other than this exception, it is always possible to construct an r-regular graph of order n, as will be demonstrated in the following theorem.

Theorem 14-6 For integers r and n, not both odd and $0 \leq r \leq n - 1$, there exists an r-regular graph of order n.

Proof: We start with the graph C_n (with the vertices as labeled in the figure below) and add to this graph to get the desired result.

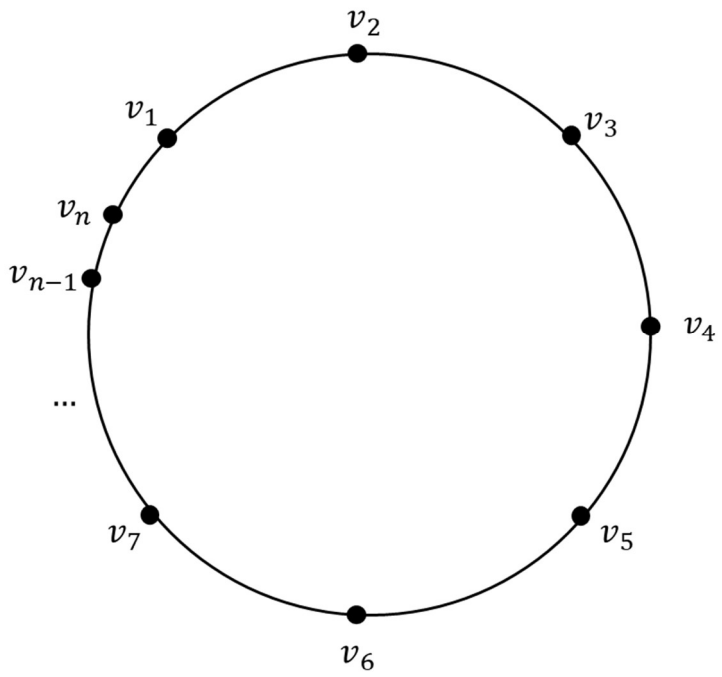


Figure 73. Graph of C_n

There are two cases, i.e., when r is even and when r is odd.

Case 1 (r is even, i.e., $r = 2k$ for some positive integer k) For each vertex v_i, draw (add) edges to the $r - 2$ nearest vertices on C_n ($k - 1$ in the clockwise direction and $k - 1$ in the counterclockwise direction). Note that $2(k - 1) = 2k - 2 = r - 2$. The updated graph (i.e., C_n plus the $r - 2$ added edges) is an r-regular graph of order n.

Case 2 (r is odd, i.e., $r = 2k + 1$ for some positive integer k) Further, we know that n must be even given the exception mentioned previously. For each vertex v_i, draw (add) edges to the $r - 3$ nearest vertices on C_n ($k - 1$ in the clockwise direction and $k - 1$ in the counterclockwise direction). Further, add edges for each pair of diametrically opposed vertices, i.e., add an edge between v_i and $v_{i+\frac{n}{2}}$. The updated graph is an r-regular graph of order n■

As an application of the above theorem, Figure 74 shows a 4-regular and a 5-regular graph of order 8. The 5-regular is obtained from the 4-regular graph by adding edges between diametrically opposed vertices (as prescribed in Case 2 of Theorem 14-6 and shown as dashed lines in the figure).

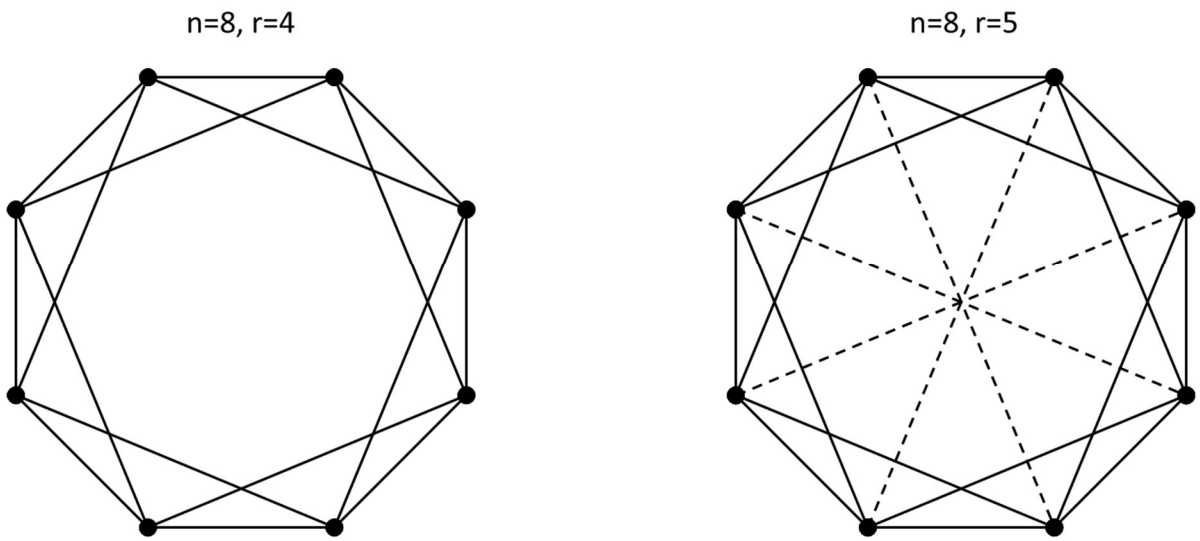


Figure 74. 4-regular and 5-regular graphs of order 8

Figure 75 depicts a 4-regular graph of order 7. This is covered by Case 1 of Theorem 14-6.

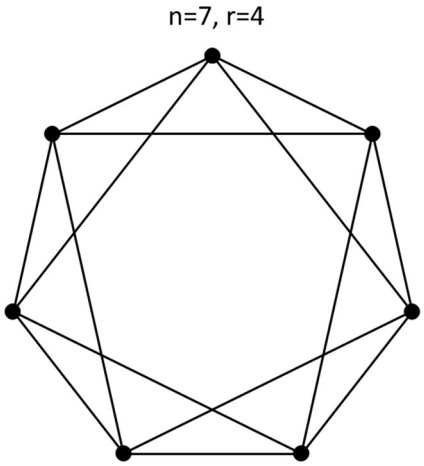


Figure 75. 4-regular graph of order 7

14.2.3 Bipartite Graphs

A **bipartite graph** is a graph whose vertices can be divided into two disjoint sets (U and V) such that every edge in the graph connects a vertex in U to a vertex in V. The vertex sets U and V are called the parts of the graph.

In terms of notation, $K_{s,t}$ is the **complete bipartite graph** of order $s + t$ with part U of order s and part V of order t. All the vertices in U are of order t and all the vertices in V are of order s. $K_{r,r}$ is an r-regular graph of order $2r$. Figure 76 shows the bipartite graphs $K_{3,2}$ and $K_{3,3}$. For each graph, one set of vertices is shown in gray (at the bottom) and the other in black (along the top).

It is also possible to have an incomplete bipartite graph where the condition of being bipartite holds but not every vertex in U has an edge with a vertex in V. For example, if one removed an edge from $K_{3,2}$, it would still be a bipartite graph but not complete.

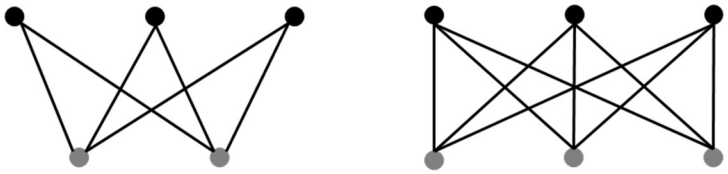


Figure 76. Examples of Bipartite Graphs

The following theorem helps to identify a graph as being bipartite or not.

Theorem 14-7 A graph G is bipartite if and only if it contains no odd cycles. ("Odd cycle" means a subgraph of the form C_{2n+1}.)

Proof: Assume G is bipartite with vertex sets U and V. Every step (edge) along the graph goes from a vertex in U to a vertex in V or from a vertex in V to a vertex in U. So, to return back to a given vertex takes an even number of steps, i.e., there are no odd cycles.

Going in the other direction, assume that every cycle in G is even.

Let x_0 be any vertex in G. For each vertex x in the same component G_1 as x_0, let $d(x)$ be the length of the shortest path from x_0 to x.

- For every vertex in G_1 whose distance from x_0 is even (i.e., even number of edges), color the vertex white. Note that x_0 is zero distance from itself (i.e., an even distance) and thus, would be colored white.
- Color the other vertices of G_1 black, i.e., vertices whose distance from x_0 is odd.

The white and black vertices (as defined above) constitute the two parts of a bipartite graph G_1, since if G_1 had an edge between two white vertices or between two black vertices, it would have an odd cycle. To see this, consider the case of two adjacent white vertices a and b. (A similar argument holds for two adjacent black vertices.) There are two possibilities:

- Case 1: If the white vertices a and b are on the same path P from x_0, the distance between a and b (along P) is even. The edge directly between a and b and the part of P between a and b constitutes an odd cycle, which contradicts our assumption. Figure 77 depicts an example cycle between a and b (shown in gray).
- Case 2: If the white vertices a and b are on different paths from x_0, again the distance between a and b is even. If we add the edge directly between a and b, we get an odd cycle, which contradicts our assumption. Figure 77 depicts an example cycle between a and b (shown in gray – basically all the edges in the diagram).

Thus, G_1 is bipartite.

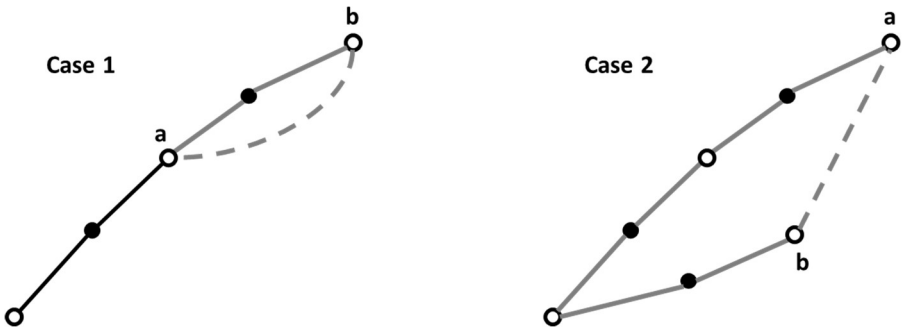


Figure 77. Odd Cycles

Do the same for each component of G.

G is comprised of one or more bipartite graphs and thus, is bipartite. The small vertices and the larger vertices are the two parts■

Using Theorem 14-7, we can determine that the graph in Figure 78 is not bipartite since it has an odd cycle (shown as dashed lines).

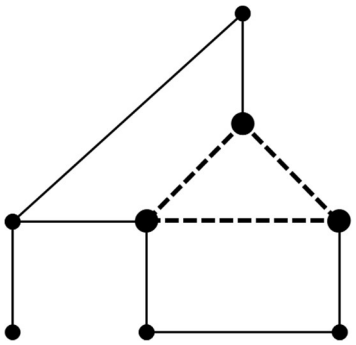


Figure 78. Non-bipartite Graph

Bipartite graphs arise in matching problems. For example, consider a set of 5 people whose first initials are A, B, C, D and E. The five people are to select from among 7 prizes at a party. However, each person is only interested in some of the prizes. The prizes are numbered from 1 to 7. The preferences of the five party goers are as follows:

- A: 1, 2, 3
- B: 1, 3, 4
- C: 1, 4, 5
- D: 1, 5, 6
- E: 1, 6, 7.

The situation can be represented by the bipartite graph shown in Figure 79. The problem is to assign (if possible) a prize to each person (from that person's list of preferred prizes). In this case, a solution is fairly easy to visualize, e.g., A gets 1, B gets 4, C gets 5, D gets 6 and E gets 7.

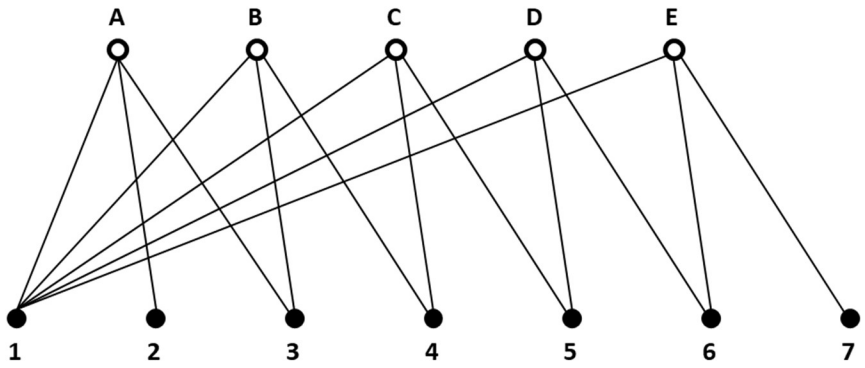


Figure 79. Prize Matching Problem

14.2.4 Trees

A **tree** is a connected graph that contains no cycles. In Figure 80, T_1 and T_2 are examples of trees.

A **forest** is a graph, all of whose connected components are trees. Graph F, in Figure 80, is a forest consisting of two components. By definition, a tree is a forest but not the other way around.

A **leaf** in a forest is a vertex of degree one.

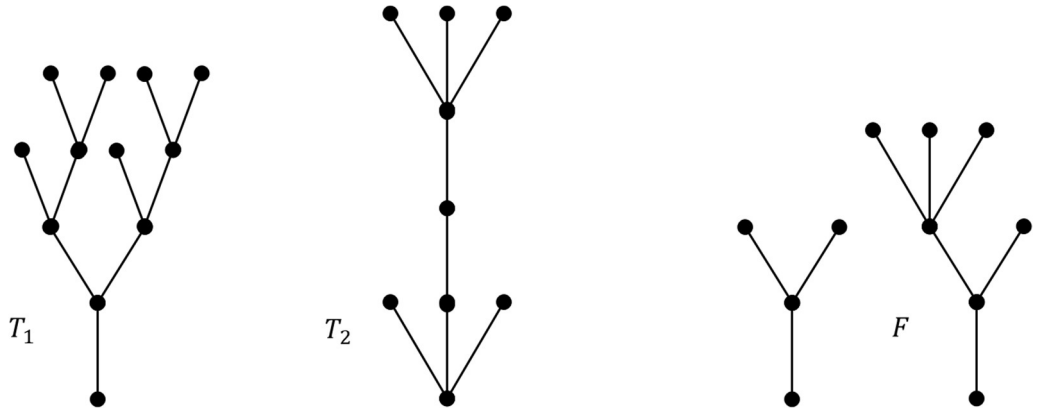


Figure 80. Example Trees and Forest

The following theorem effectively provides an alternate definition of a tree.

Theorem 14-8 A graph G is a tree if and only if every two vertices are connected by exactly one path.

Proof: If every pair of vertices in G are connected by a single path, then G is connected. If G contained a cycle C, then any two vertices on C could be connected by two paths which contradicts the initial assumption. So, G must have no cycles and by definition, G is a tree.

Going in the other direction, assume G is a tree. By definition, we know that G is connected and thus, every two vertices have at least one path between them. Assume two vertices u and v are connected by two different paths P and Q. The proof is complicated by the fact that P and Q may partially overlap. However, starting from u, there must be a first vertex x on P and Q such that the next vertex on P and Q (leading to v) is different. Further, since P and Q both terminate on v, there

must be a first vertex y after x (leading to v) that is on both P and Q. (Note that x and y cannot be the same vertex, since P and Q would then be the same path.) The paths from x to y on P and Q have no vertices in common and have the same starting and end vertex, and thus, form a cycle but this contradicts the initial assumption that G is a tree. So, u and v (which we selected arbitrarily) are only connected by a single path■

The next two theorems concern properties of leaves with regard to trees.

Theorem 14-9 If a tree T has at least two vertices, then it has at least two leaves.

Proof: For a tree with two vertices, there are exactly two leaves.

For the case where there are 3 or more vertices, let P be a path of maximum length in T. Represent P as the ordered list $(v_1, v_2, \dots, v_n)$. If v_1 is adjacent to any other vertex in P, this would create a cycle which is not possible since we are given that T is a tree. So, v_1 is not adjacent to any other vertex in P. If v_1 were adjacent to a vertex v not in P, then we would get a longer path than P, i.e., $(v, v_1, v_2, \dots, v_n)$ which is a contradiction to the assumption about P. Thus, v_1 is only adjacent to v_2 and so, v_1 is (by definition) a leaf. Similarly, v_n is a leaf■

Theorem 14-10 If a tree T has n vertices, then it has n-1 edges.

Proof: The proof is by induction.

For $n = 1$, we have a tree consisting of a single vertex and thus $n - 1 = 0$ edges.

Assume the theorem is true for $n = k$, i.e., if a tree has k vertices then it has $k - 1$ edges.

Consider any tree T with $k + 1$ vertices. By Theorem 14-9, T has at least two leaves. Remove one leaf and the single edge attached to the leaf. This leaves us with T' (a tree with k vertices). By the induction assumption, T' has $k - 1$ edges. Thus, T has k edges■

14.2.5 Subgraphs

A graph H is called a **subgraph** of a graph G if every vertex of H is a vertex of G, and every edge of H is an edge of G.

If H is a subgraph of G having the same vertices as G (but not necessarily the same edges), then H is a **spanning subgraph** of G.

In Figure 81, H and K are subgraphs of G. Further, K is a spanning subgraph of G, and H is not a spanning subgraph of G.

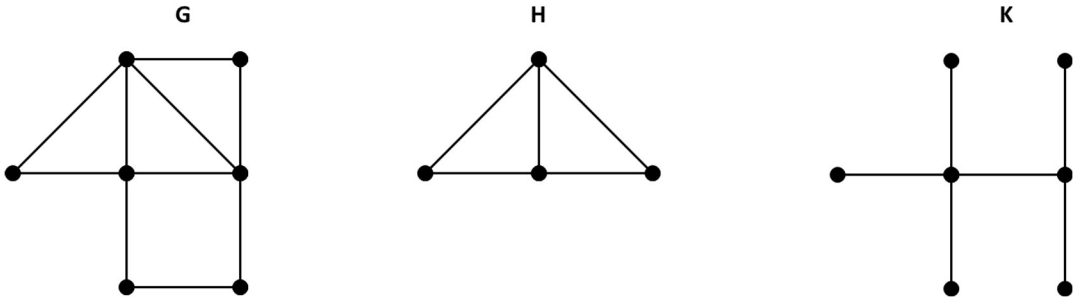


Figure 81. Examples of Subgraphs

14.2.6 Isomorphic Graphs

The same graph can be drawn in multiple ways and in many cases, it is far from obvious that the graphs are the same structurally, e.g., Figure 82 depicts three different renderings of the same graph (known as the Peterson graph).

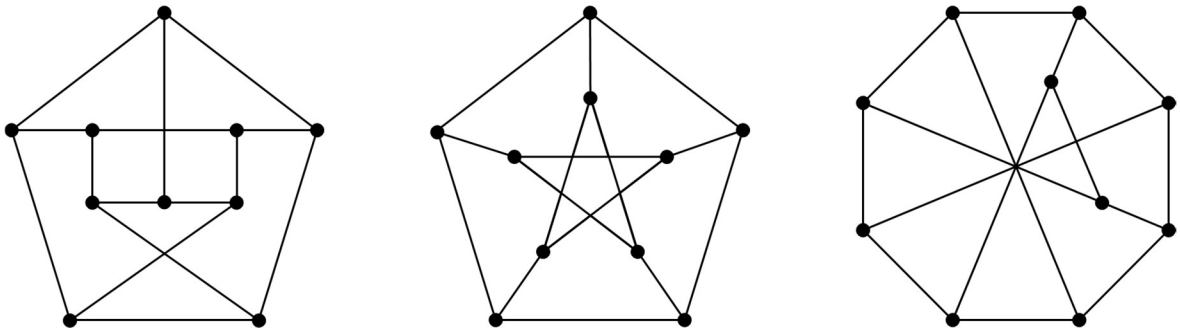


Figure 82. Isomorphic Renderings of the Peterson Graph

Two graphs are said to be **isomorphic** if they have the same structure. In other words, graphs G and H are isomorphic graphs if the vertices of G can be relabeled to produce H. More formally, graphs G and H are isomorphic if one can define a bijection between the vertex sets of G and H, i.e., $f: V(G) \longrightarrow V(H)$, such that any two vertices u and v of G are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in H.

Figure 83 shows a bijective mapping f between two isomorphic graphs.

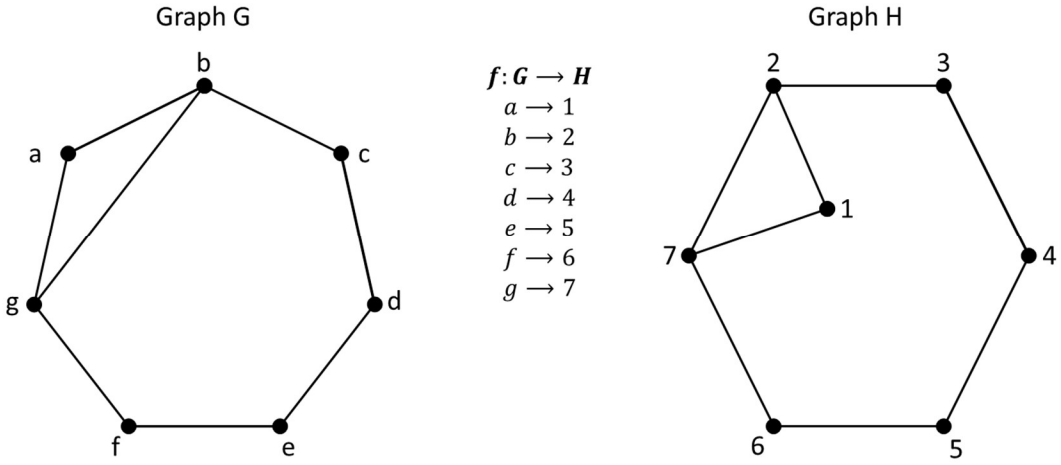


Figure 83. Bijective Mapping between Isomorphic Graphs

14.3 Connectivity

14.3.1 Trails, Path and Cycles

Connected graphs were introduced in Section 14.1. In this section, graph connectivity is explored in more detail.

A **trail** between vertices u and v is a subgraph in a graph G consisting of edges (and associated vertices) that lead from u to v. Vertices can be repeated in a trail but edges may not. A trail is

written as an ordered sequence of vertices where there is an edge between each adjacent vertex in the sequence. If $u = v$ (i.e., start and end vertices are the same), then the trail is said to be closed; otherwise, the trail is classified as open. A closed trail is also referred to as **circuit**.

A **path** is a trail with the added restriction that vertices may not be repeated (other than the two endpoints, which may be the same).

If the endpoints of a path are the same, the path is a **cycle**.

As depicted in the Venn diagram below (Figure 84), all cycles are paths and all paths are trails, but not all trails are paths, and not all paths are cycles. All cycles are circuits (not explicitly shown in the figure).

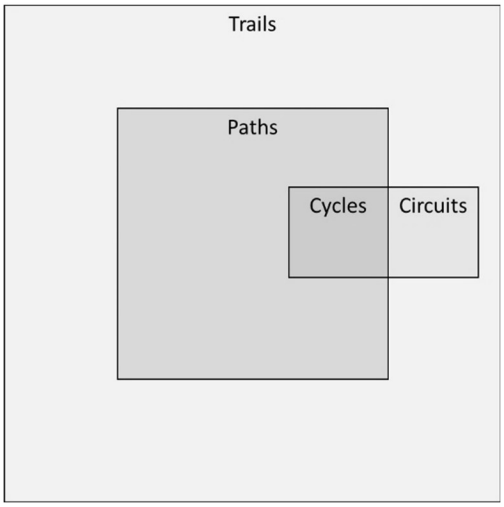


Figure 84. Relationship between Trail, Path and Cycle

A circuit (i.e., closed trail) in a connected graph G that contains every edge of G is called an **Eulerian circuit** (named after the famous mathematician Leonhard Euler). An open trail that contains every edge of G is called an **Eulerian trail**. A connected graph is classified as an **Eulerian graph** if it contains an Eulerian circuit.

The **length** of a trail is the number of edges in the trail.

The following are trails within the graph shown in Figure 85:

- $P = (a, b, f, c)$ is a path from vertex a to c along edges ab, bf and fc. The length of P is 3.
- $C_1 = (a, e, f, b, a)$ is a cycle of length 4 and by definition, also a circuit.
- $T = (g, h, e, a, d, h, i)$ is not a path because the interior vertex h is repeated. However, T is a valid trail of length 6.
- $C_2 = (d, a, e, h, g, e, d)$ is a circuit but not a cycle since the vertex e is repeated in the interior of the trail.
- Graph G is not Eulerian for reasons that will be explained in Section 14.5.

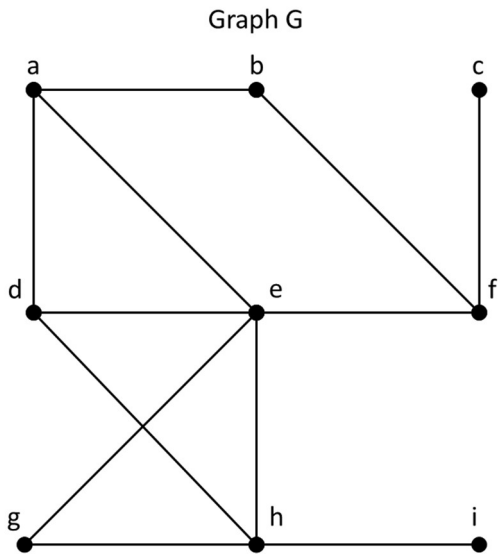


Figure 85. Paths and Distances

14.3.2 Cut-vertices and Bridges

A **bridge** (sometimes called a cut-edge) is an edge of a graph whose deletion increases the graph's number of connected components. Equivalently, an edge is a bridge if and only if it is not contained in any cycle. The graph in Figure 85 is bridgeless. The edge cd in Figure 86 is a bridge.

A vertex is a **cut-vertex** of graph G if $G - v$ has more components than G, where $G - v$ is defined to be the graph G with vertex v and all edges adjacent to v removed. In other words, v is a cut-vertex in a connected graph G if $G - v$ is disconnected. In Figure 86, b, c and d are cut-vertices.

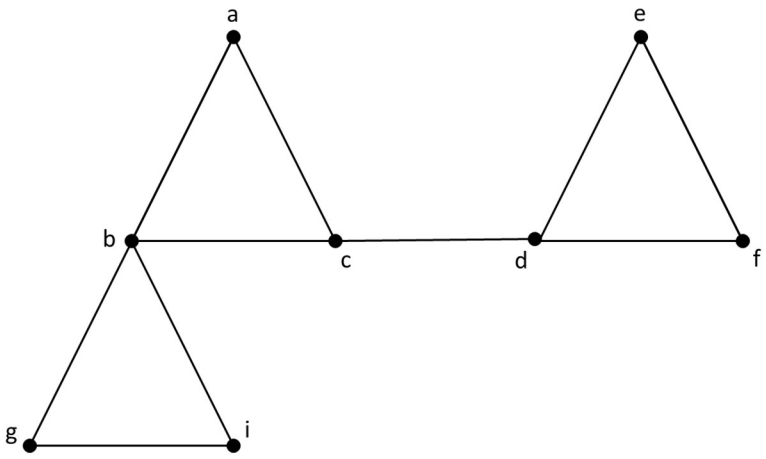


Figure 86. Bridges and Cut-Vertices

Figure 87 provides an example of "subtracting" a vertex from a graph. In particular, Figure 87 depicts $G - e$ where G is the graph from Figure 85. In this case, e is not a cut-vertex since $G - e$ has the same number of components as G, i.e., just one.

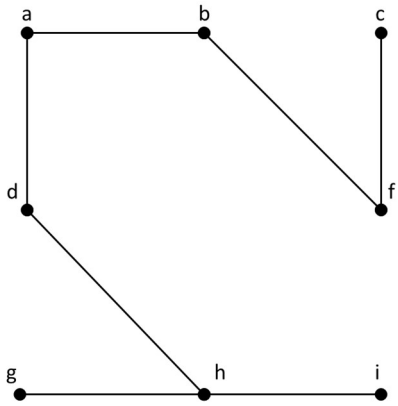


Figure 87. Example of G minus a vertex

14.4 Minimum Spanning Trees

Thus far, the edges in all the graphs presented in this book have had equal weight. Recall from the discussion on trail length that we counted each edge as 1 unit length. In many applications, the edges are not of equal weight, e.g., the length of roads between two points or the cost of a transmission line between two switching elements in a telecommunications network.

Figure 88 depicts a weighted graph, with the various weights indicated next to each edge. For example, the weight of edge bd is 7. In terms of notation, we write this as $w(bd) = 7$.

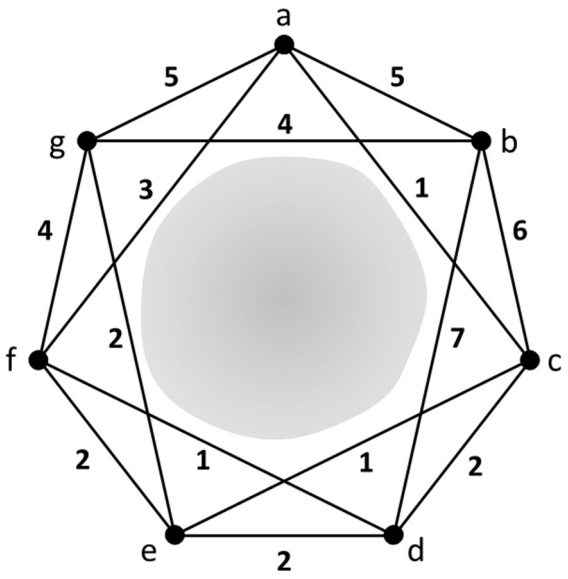


Figure 88. Weighted Graph

Assume the vertices in Figure 88 are planned sites for telecommunications equipment (sometimes referred to as Central Offices) and we want to provide minimum cost connectivity among the sites. The gray thing in the center of the figure is meant to represent water and due to cost considerations, paths crossing the water (perhaps underground cables) are excluded from

consideration. So, for example, we exclude the possibility of an edge between vertex f and vertex c and other similar edges that involve crossing the water obstacle (or taking a very long route).

For the problem at hand, we want to determine a minimum spanning tree of graph G. In general, a **minimum spanning tree** is a subgraph T of a connected, edge-weighted undirected graph such that T is a tree with the minimum possible total edge weight. It is possible for a graph to have several minimum spanning trees. The minimum spanning tree problem was first posed by Otakar Borůvka in 1926 concerning the least cost layout for a power-line network. In what follows, we will discuss a solution to the problem attributed to Joseph Bernard Kruskal. The solution is known as Kruskal's algorithm.

Kruskal's algorithm is summarized below. Assume the algorithm is applied to a graph G of order n.

1. Sort all the edges in order of their weight.
2. Start with an empty spanning tree S (i.e., no vertices or edges).
3. Pick an edge x of smallest weight from the edges not yet in S and which have not been previously discarded (and placed into set D, as defined below). If there is a tie in terms of weight, arbitrarily pick one. If the addition of x to S does not form a cycle, update S to include x. Else, discard x. Keep track of discarded edges in the set D.
4. Repeat Step #3 until there are $n - 1$ edges in S.

Figure 89 depicts the application of Kruskal's algorithm to the graph from Figure 88. As edges are added to the spanning tree, they are shown in heavy dashed lines. The algorithm goes smoothly (just selecting the lowest weight edge and breaking ties) until the fifth iteration (middle graph in bottom row). At this point, adding edge ed, ef or fa will create a cycle in the spanning graph. The next possibilities are gb and gf (both of weight 4) but gf leads to a cycle, and so, we add gb. This is the last step since we now have $n - 1 = 6$ edges in the spanning tree.

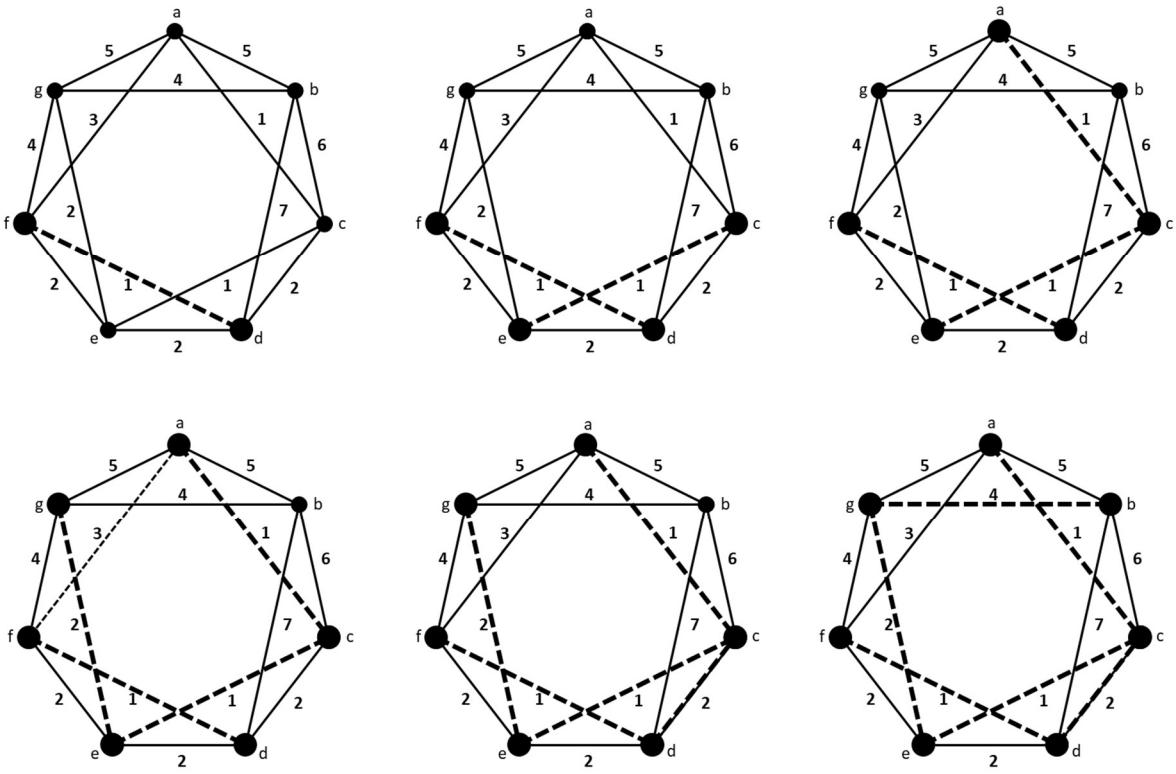


Figure 89. Example of Kruskal's Algorithm

The spanning tree in the last step of Figure 89 is a bit difficult to visualize as a tree. In Figure 90, the minimum spanning tree (with some rearrangement) is shown in isolation from the rest of the graph.

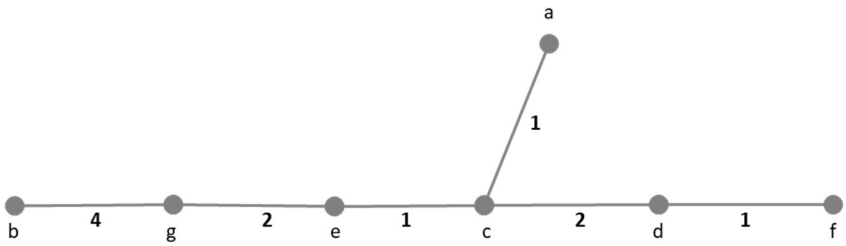


Figure 90. Minimum Spanning Tree

14.5 Traversing Graphs

One would be hard pressed to find an introductory graph theory book that does not mention the Königsberg bridge problem. The story goes as follows:

> In the town of Königsberg (now Kaliningrad, Russia) many of the townsfolk spent Sunday afternoon walking about the town. The question arose as to whether it was possible to visit all four regions of the town (shown as W, X, Y and Z in Figure 91) while crossing each of the seven bridges (a, b, c, d, e, f and g) exactly once.

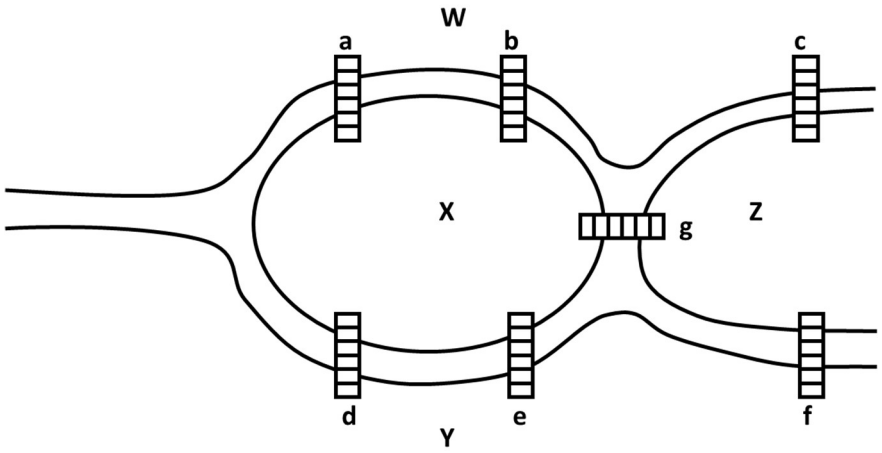


Figure 91. Königsberg Bridge Problem

The first step in solving the Königsberg bridge problem is to represent the situation as a graph (as shown in Figure 92). This is a multigraph because of the multiple edges between X and W, and between X and Y.

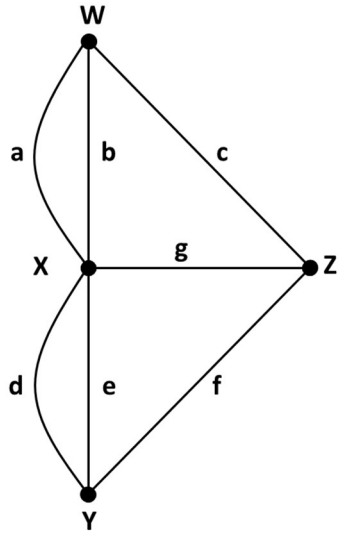


Figure 92. Graphical Representation of Königsberg Bridge Problem

The following theorem (stated here without proof) allows one to determine which graphs are Eulerian graphs.

Theorem 14-11 A connected graph G (either a simple graph or a multigraph) is Eulerian (i.e., contains a closed trail that includes every edge of the graph) if and only if the degree of every vertex of G is even.

For the problem at hand, the above theorem implies that the Königsberg is not Eulerian since, in fact, none of the vertices is even. So, it is not possible to visit all the regions in the town without crossing each bridge just once.

The following theorem (also stated without proof) follows from Theorem 14-11.

Theorem 14-12 A connected graph (either a simple graph or a multigraph) contains an Eulerian trail (i.e., an open trail that includes every edge of the graph) if and only if exactly two of its vertices have odd degree. In this case, the two vertices of odd degree are at opposite ends of the Eulerian trail.

14.6 Exercises

1. Show all the graphs (of order 4) for the case $n = 3$ in Theorem 14-5. ***Hint**: There are 4 such graphs.*
2. Regarding Theorem 14-5, how many graphs are there for the case *n=5*? ***Hint**: In general, the answer is a power of 2.*
3. Draw K_4 without any of the edges crossing each other. Is it possible to draw K_5 without crossing any edges? ***Hint**: See the Wikipedia article entitled "Planar graph" [71].*
4. How many edges are there in C_n? How many edges are there in K_n?
5. Is there a 2-regular graph of order 6 that is not a single cycle? ***Hint**: Try two cycles.*
6. Draw a 4-regular graph of order 6. Draw a 3-regular graph of order 10.
7. Prove that if a tree T is of order n and size m, then the sum of the degrees of the vertices of T is $2n - 2$. ***Hint**: Use Theorem 14-1 and Theorem 14-10.*
8. Are graphs G and H in Figure 93 isomorphic? If so, provide a structure-preserving bijective mapping between the vertices of G and H? ***Hint**: the vertex labeling in the figure is suggestive of the solution.*

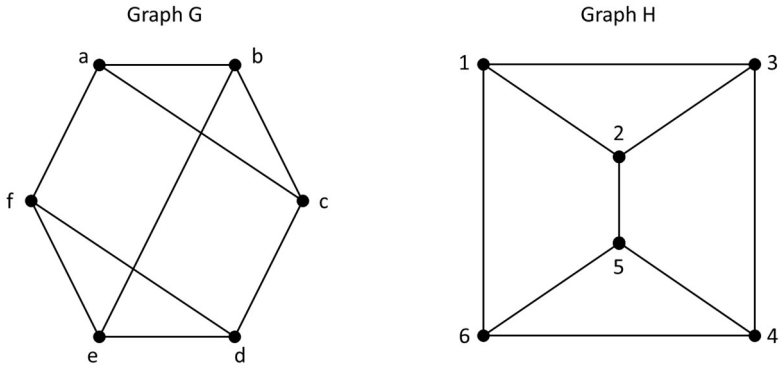


Figure 93. Are graphs G and H isomorphic?

9. In the 4th step of the application of Kruskal's algorithm in Figure 89 (bottom left graph), choose to add edge ed rather than eg, and then complete the algorithm.
10. Find the Eulerian trail in Figure 86. ***Hint**: Use* Theorem 14-12 *to find the ends of the Eulerian trail*.

11. Show that the graph in Figure 94 has an Eulerian trail.

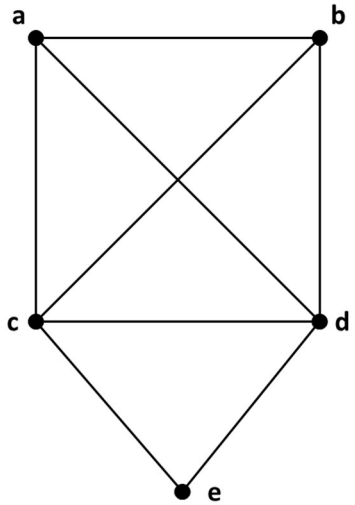


Figure 94. Eulerian Graph

15 Linear Algebra

Linear algebra entails the study of **vectors** (a quantity, such as velocity, completely specified by a magnitude and a direction) and **matrices** (a rectangular array of numbers for which operations such as addition and multiplication are defined), and the transformation of these entities. More generally, linear algebra is an area of study in mathematics that concerns itself primarily with the study of vector spaces and the linear transformations of vector spaces.

The uses of linear algebra are extensive. The following is a brief list:

- Ranking in search engines: For example, the original ranking algorithm for Google used linear algebra to rank which web pages should be shown first.
- Linear programming is a method to achieve the best outcome (such as maximum profit or lowest cost) in a mathematical model whose requirements are represented by linear relationships. Linear algebra is at the foundation of linear programming.
- Error correcting codes: The use of linear algebra in coding theory is extensive. The problem is to encode data in such a way that if the encoded data is altered, it is still possible to recover the unencoded data. Such schemes are called error correcting codes, and the simplest ones encode data as vectors in a vector space (concepts to be discussed in this section).
- Computer graphics: A key part of graphics is projecting a three-dimensional scene onto a two-dimensional screen. Projections can be modeled as linear mappings (a concept to be discussed in this section). Further, rotations, scaling, and perspective are all implemented and analyzed using linear algebra.
- Linear Algebra is a fundamental building block for data science, pattern recognition, and Machine Learning (ML).

In what follows, vectors and matrices are studied first. This is followed by the introduction of vector spaces of which vectors and matrices are examples.

15.1 Matrices and Vectors

As noted, a matrix is a rectangular array of numbers. Figure 95 depicts a matrix with 3 rows (horizontal lists of numbers) and 6 columns (vertical lists of numbers). The dimension (or alternately, “size”) of the matrix is 3x6 (with the numbers of rows coming first, followed by the number of columns). Each entry in a matrix is identified by its row and column position. For example (in Figure 95), -7 is the entry at position (2,3) and .125 is at position (3,4).

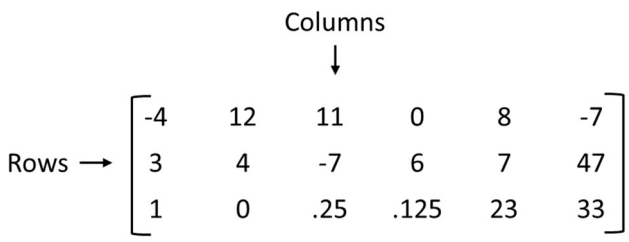


Figure 95. Example of a 3x6 Matrix

The top part of Figure 96 shows a matrix being multiplied by a constant (referred to as a scalar in this case). This type of multiplication, where each entry in a matrix is multiplied by the same number, is known as **scalar multiplication**. The matrix is scaled and thus the term scalar multiplication.

The bottom part of Figure 96 shows an example of **matrix addition**. This is a straightforward process where entries in the same position are added and then put into the resulting matrix.

$$2\begin{bmatrix} 1 & 2 \\ 13 & 3 \\ -6 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 26 & 6 \\ -12 & 8 \end{bmatrix}$$

$$\begin{bmatrix} -4 & 12 \\ 3 & 4 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 8 & 3 \\ -3 & 5 \\ -9 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 15 \\ 0 & 9 \\ -8 & 8 \end{bmatrix}$$

Figure 96. Scalar Multiplication and Addition of Matrices

It is possible to combine scalar multiplication and addition. For example, assume that A and B are both matrices of size $m\ x\ n$, then $C = 3A - 7B$ is another matrix of size m x n. Matrix C is formed by multiplying A by 3, multiplying B by -7 and then adding the result to form C.

A vector is simply a matrix consisting of either one row or one column. Figure 97 shows two essentially the same vectors (the only difference is their dimension). Vectors can be added and multiplied by a scalar just like any other matrix. There is also a geometric interpretation for vectors (see the next section).

$$\begin{bmatrix} 5 \\ -3 \\ 7 \end{bmatrix}$$

3x1 vector

$$[\ 5\ \ -3\ \ 7\ ]$$

1x3 vector

Figure 97. Example Vectors

Matrix multiplication is not straightforward nor obvious as to why it is defined as it is. By way of motivation, consider the equation $5x - 3y + 7z = 9$. This equation can be represented by the multiplication of the matrices shown in Figure 98. The single row of the matrix on the left is multiplied times the adjacent single column such that

- 5 is multiplied times x
- -3 is multiplied times y
- 7 is multiplied times z
- the above three terms are added together (this is the left-hand side of the equation)
- the right-hand side of the equation is represented by the 1x1 matrix 12.

$$[\,5\ \ -3\ \ 7\,]\begin{bmatrix} x \\ y \\ z \end{bmatrix} = [\,12\,]$$

Figure 98. Representation of an Equation via Matrix Multiplication

In general, the multiplication of two vectors in the manner depicted above is known as the **dot product** of the vectors. The general formula for the dot product is shown in Figure 99.

$$[\,x_1\ x_2\ \ldots x_n\,]\begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

Figure 99. Dot Product of two Vectors

As another example of matrix multiplication, consider the system of equations and the associated matrix representation shown in Figure 100. In a manner similar to the previous example, the dot product of each row of A is taken with the column matrix X. The figure highlights (with ellipses) the dot product of row 2 of A with column 1 of X to get $2x_1 - 7x_2 + 8x_3$ which is then equated to the 2nd row of Y, i.e., the number 6 in this case.

$$\begin{aligned} x_1 + 3x_2 + 7x_3 &= 4 \\ 2x_1 - 7x_2 + 8x_3 &= 6 \\ -3x_1 + 8x_2 - 5x_3 &= -4 \end{aligned}$$

$$\underset{\text{A}}{\begin{bmatrix} 1 & 3 & 7 \\ 2 & -7 & 8 \\ -3 & 8 & -5 \end{bmatrix}} \underset{\text{X}}{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} = \underset{\text{Y}}{\begin{bmatrix} 4 \\ 6 \\ -4 \end{bmatrix}}$$

Figure 100. Representation of a System of Equations via Matrix Multiplication

In general, matrices A and B are **multiplied** to get matrix C by taking the dot product of the i^{th} row of A with the j^{th} column of B to get the entry in row i and column j of C. In order for matrix multiplication to be possible (given the above definition), matrix A must have the same number of columns as the number of rows in B. So, if A is an $n\ x\ m$ matrix and B is an $m\ x\ p$ matrix, $C = AB$ is an $n\ x\ p$ matrix.

Consider the example shown in Figure 101. To get the first row of C, we take that dot product of the first row of A with each of the columns of B. For example, the (1,3) entry in C is the dot product of the first row of A with the third column of B, i.e., $1 \cdot 3\ +\ 3 \cdot 13\ =\ 42$. The second row of C is gotten by successively taking the dot product of the second row of A with each column of B. Finally, the third row of C is obtained by taking the dot product of the third row of A with each column of B. Notice that BA is undefined since the number of columns in B does not equal the number of rows in A.

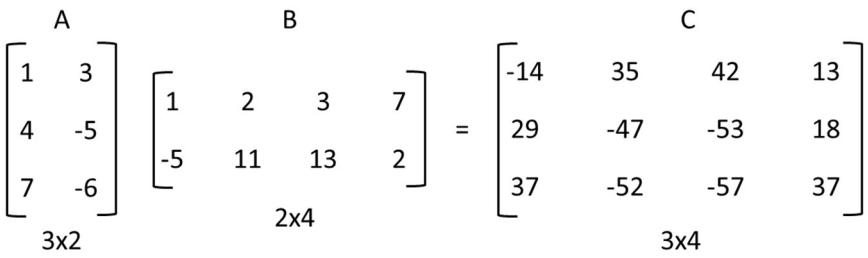


Figure 101. Matrix Multiplication

Now that multiplication of matrices is defined, a somewhat obvious question to ask is whether there is some sort of identity matrix I such that I multiplied times another matrix A is the matrix A, i.e., $IA = A$. In fact, there are many such matrices, i.e., one for each positive integer n. The structure is very simple, i.e., 1s on the diagonal and 0s everywhere else. Figure 102 shows the identity matrix for $n = 3$. If clear from the context, we just write I for the identity matrix; otherwise, a subscript can be appended to make clear the size of the identity matrix, i.e., I_n. If A is an m x n matrix, then $I_m A = AI_n = A$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Figure 102. Identity Matrix of Size 3 x 3

It turns out that some matrices have inverses where A^{-1} is the **inverse** of matrix A if $A^{-1}A = AA^{-1} = I$. Consider again the example in Figure 100. If an inverse for A can be found, it can then be multiplied on both sides of the equation $AX = Y$ to get a solution to the problem, i.e., $A^{-1}AX = IX = \boldsymbol{X = A^{-1}Y}$. Herein lies a key benefit in defining matrix multiplication as we have above, i.e., it allows for the representation and solution of systems of linear equations. The Gauss-Jordan elimination (described in Section 15.3.2) provides a technique to determine the inverse of a matrix, which, in turn, can be used to solve a system of equations.

The transpose of a matrix A (written as A^t) maps each row of A into a column. This has the effect of transposing the indices of A to get A^t, i.e., the (i, j) entry of A is the (j, i) entry of A^t. Figure 103 shows a matrix and its transpose. Clearly, $(A^t)^t = A$.

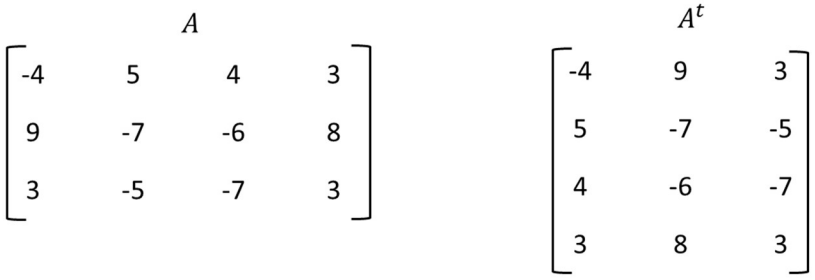


Figure 103. Transpose of a Matrix

15.2 Vectors – Geometric Approach

Vectors also have a geometric interpretation in 1, 2 and 3 dimensional real space (represented as $\mathbb{R}, \mathbb{R}^2$ and $\mathbb{R}^3$, respectively. When discussing the geometric aspects of vectors in this book, we will

not make a distinction between a vector represented as a row or as a column matrix, and just use the simplified notation of an ordered sequence, e.g., $(4, -3,9) \in \mathbb{R}^3$. The use of this geometric interpretation is prevalent in physics and geometry.

In Figure 104, there are three vectors, i.e., $\vec{u} = \overrightarrow{AB} = (4,2)$, $\vec{w} = \overrightarrow{DE} = (4,2)$ and $\vec{v} = \overrightarrow{AC} = (-2,4)$. The first number in the tuple indicates the distance from the tail end (non-arrow) to the arrow end of the vector in the horizontal direction, and the second number is the distance from the tail end to the arrow end of the vector in the vertical direction. Notice that there are two ways to reference a vector, i.e., either via a label such as $\vec{u}$ or using the start (tail) and end (arrow) points such as $\overrightarrow{AB}$.

Recall the earlier definition of a vector, i.e., a quantity completely specified by a magnitude and a direction. This means that vectors $\vec{u}$ and $\vec{w}$ (in Figure 104) are considered to be the same, but at various locations in $\mathbb{R}^2$.

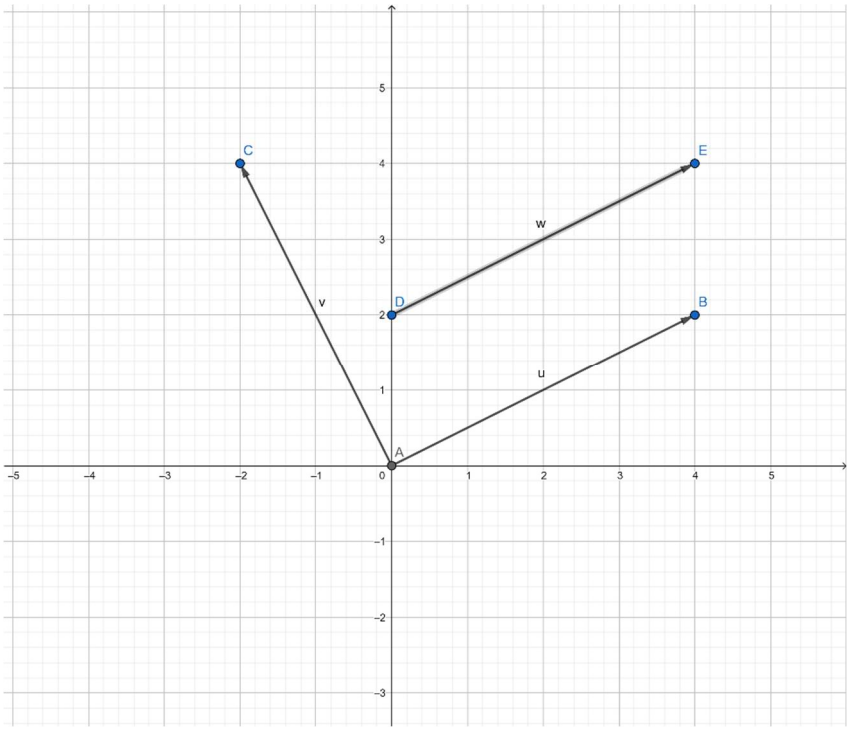


Figure 104. Vector Examples

The geometric approach to vector addition is to place the tail of one vector at the arrow end of the other, and then draw a vector from the tail of the first vector to the arrow end of the second vector. Figure 105 depicts the addition of vectors $\vec{u}$ and $\vec{v}$ (see the vector $\vec{u} + \vec{v} = \overrightarrow{AD}$). If we use matrix addition (as defined in the previous section), then we get $\vec{u} + \vec{v} = (4,2) + (-2,4) = (2,6)$ which, as can be seen from the figure, is the same as $\overrightarrow{AD}$.

The geometric approach for subtraction is very similar to addition. First, create the negative of the vector to be subtracted (this is just a vector of the same length but in the opposite direction) and then add. Figure 105 shows $-\vec{v} + \vec{u} = \overrightarrow{CB}$ (which is the same as $\vec{u} - \vec{v}$).

Similar geometric interpretations for vector addition and subtraction are possible in $\mathbb{R}^3$, or for that matter in $\mathbb{R}$ (although not very interesting in this case).

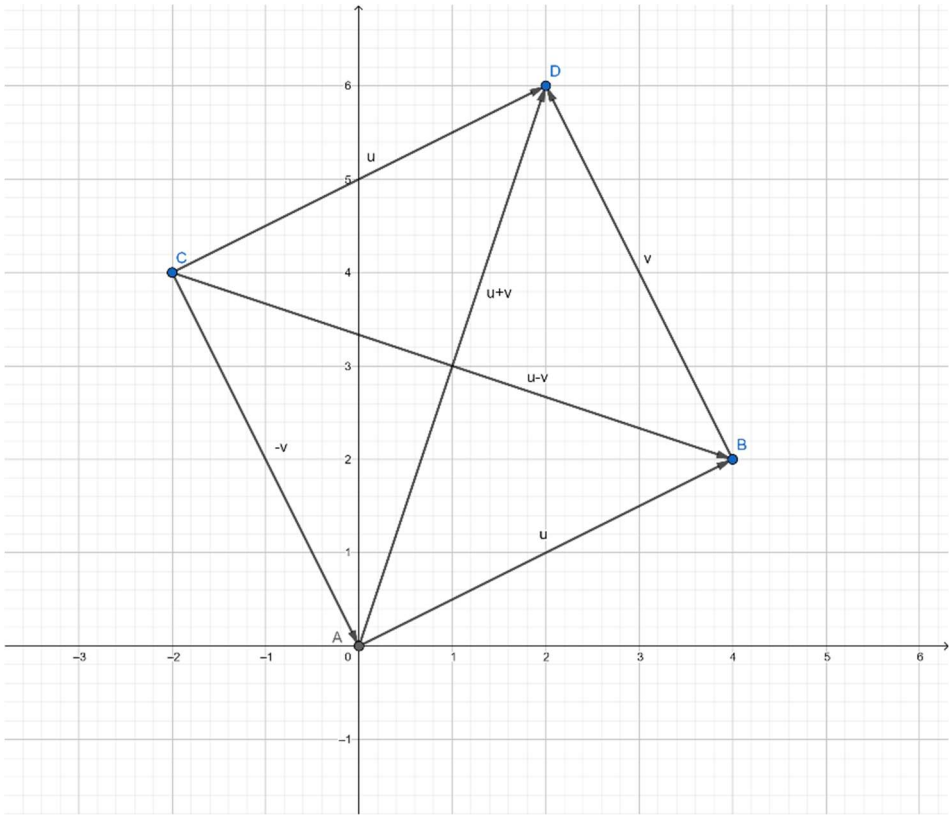


Figure 105. Vector Addition and Subtraction

Consider the line with equation $y = 2x + 3$. The line can be parameterized by the equations

$$x = t$$

$$y = 2t + 3$$

Form the vector $\vec{w} = (t, 2t + 3) = t(1,2) + (0,3)$. This is an equivalent representation of the line $y = 2x + 3$. As shown in Figure 106, the line $y = 2x + 3$ can be written as the vector $\vec{v} = (0,3)$ plus scalar multiples of the vector $\vec{u} = (1,2)$. For example, if we take $t = 2$, we get the point $D(2,7)$ which is on the line, as shown in the figure.

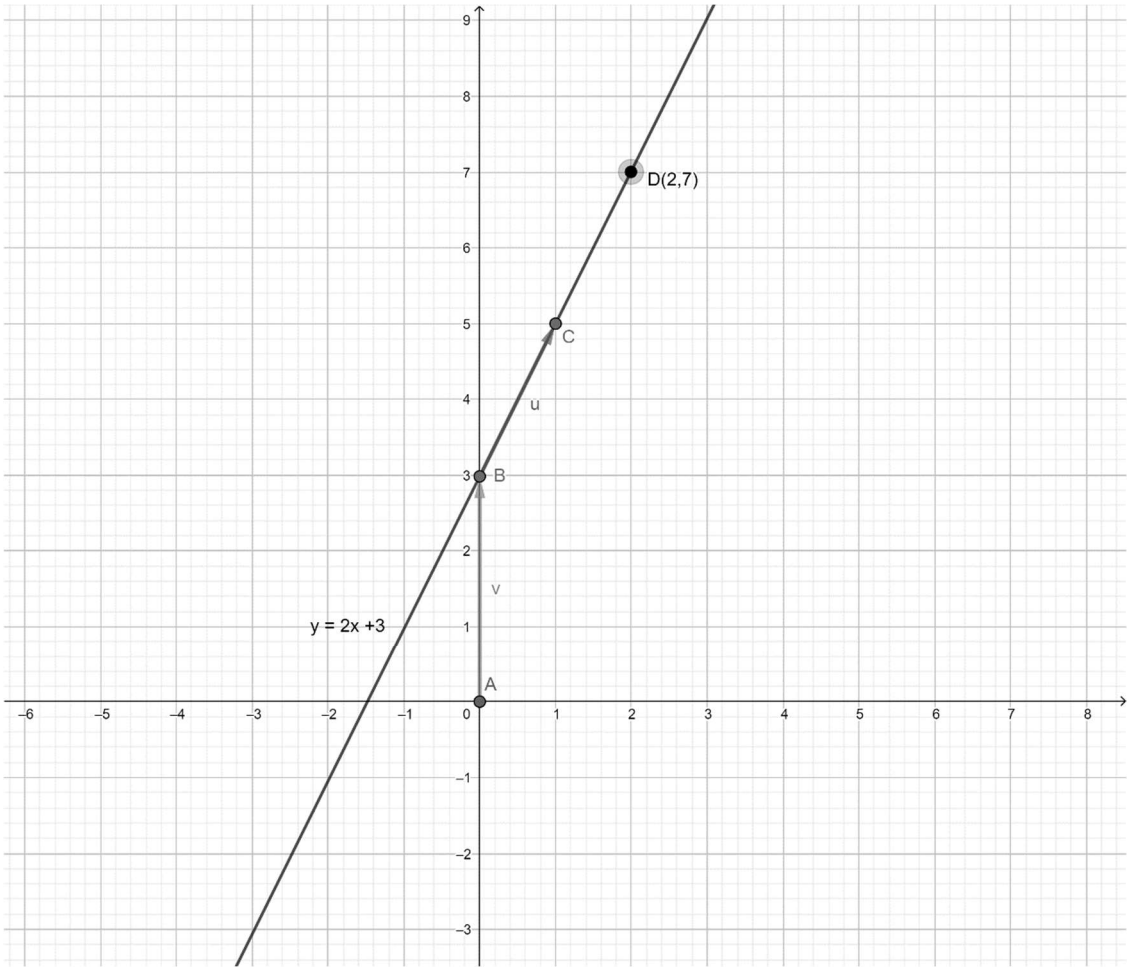


Figure 106. Vector Representation of a Line

15.3 Systems of Linear Equations

We've already seen how to represent a system of equations in terms of matrices. This section covers an algorithm that can be used to solve systems of equations.

15.3.1 Geometric Considerations

For a system of two linear equations (i.e., two lines in a plane), there are three possibilities, i.e., the lines coincide (infinitely many solutions), the lines cross at one point (one solution) or the lines are parallel (no solutions).

An equation with three variables, e.g., $2x - 2y + z = 3$, is a plane. To see this, we parameterize the equation, i.e.,

$$\text{let } x = t$$

$$\text{and } y = s$$

$$\text{which implies } z = 3 - 2t + 2s$$

Thus, all solutions to the equation are of the form

$$(t, s, 3 - 2t + 2s) = (0,0,3) + t(1,0,-2) + s(0,1,2).$$

If we let $B = (0,0,3)$, $C = (0,1,2)$ and $A = (1,0,-2)$, then the vectors $\overrightarrow{BC}$ and $\overrightarrow{BA}$ define a plane going through the point $(0,0,3)$, see the vertical plane in Figure 107. The horizontal plane in the figure is the xy-plane.

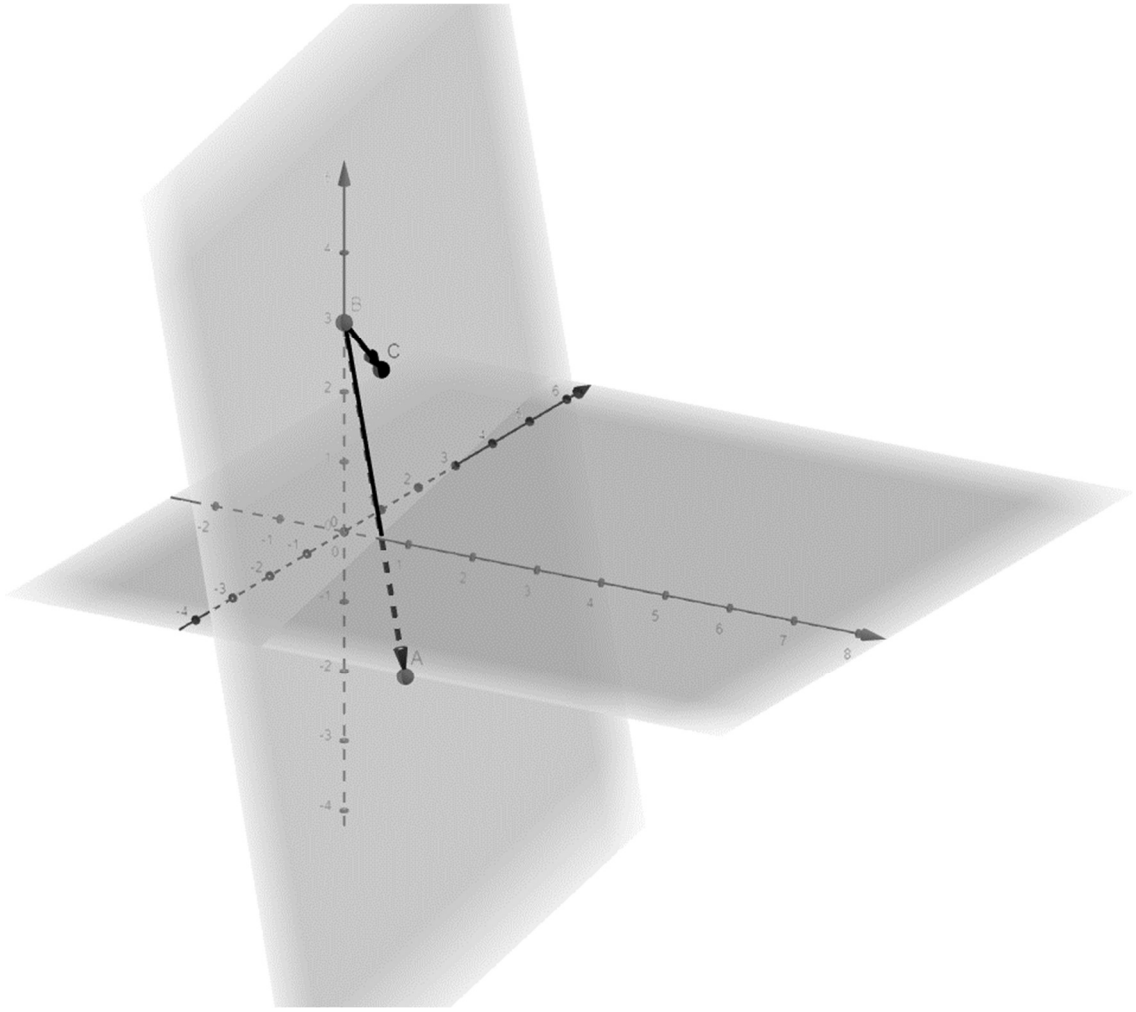


Figure 107. Vectors Defining a Plane

If we have three equations with three variables (i.e., three planes), then we have the following possibilities:

- The planes coincide (infinite number of solutions)
- At least two of the planes are parallel (no solutions)
- Two planes coincide and the other plane intersects the two incident planes in a line (infinite number of solutions)
- All three planes intersect at one point (unique solution).

In general, each linear equation of n variables determines a **hyperplane** of dimension n-1. The possible solution sets are as follows:

- The hyperplanes coincide (infinite number of solutions where the solution set is of dimension $n-1$)
- At least two of the hyperplanes have no common points (no solutions)
- Various combinations of the hyperplanes coincide (no solutions or an infinite number of solutions of dimension $n-2$ or less)
- All the hyperplanes intersect at one point (unique solution).

15.3.2 Gaussian Elimination

Gaussian elimination is an algorithm for solving a system of linear equations. It entails a sequence of operations performed on the matrix of coefficients corresponding to the linear equations. The procedure is named after mathematician Carl Friedrich Gauss (1777–1855) but some special cases of the method were known to Chinese mathematicians as early as circa 179 AD.

The algorithm is based on three types of row operations on the matrix representing a system of equations. The critical point here is that the following row operations do not change the solution to the system of equations:

- Swap the order of two rows
- Multiply a row by a constant
- Add a multiple of one row to another.

Table 52 shows an example of the Gaussian elimination. The equation format is not typically used and is added here to help the reader understand the example. The matrix format on the right is equivalent to the equation format. The middle column of the table shows the matrix row operations in going from one step to the next, e.g., $R_2 - 2R_1$ means replace the existing row 2 with row 2 minus 2 times row 1. The idea is to get the matrix in a form where one can just read off the solution, as we have done in this particular example.

Table 52. Gaussian Elimination – Single Solution

Equation Format	Row Operations	Matrix Format
$\boldsymbol{x+y=1}$ $\boldsymbol{2x+3y-z=0}$ $\boldsymbol{-3x-3y+z=1}$		$\left\langle \begin{matrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ -3 & -3 & 1 \end{matrix} \middle\| \begin{matrix} 1 \\ 0 \\ 1 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y-z=-2}$ $\boldsymbol{z=4}$	$R_2 - 2R_1$ $R_3 + 3R_1$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{matrix} \middle\| \begin{matrix} 1 \\ -2 \\ 4 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y=2}$ $\boldsymbol{z=4}$	$R_2 + R_3$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \middle\| \begin{matrix} 1 \\ 2 \\ 4 \end{matrix} \right\rangle$
$\boldsymbol{x=-1}$ $\boldsymbol{y=2}$ $\boldsymbol{z=4}$	$R_1 - R_2$	$\left\langle \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \middle\| \begin{matrix} -1 \\ 2 \\ 4 \end{matrix} \right\rangle$

There is not always a single solution, e.g., consider the linear system of equations in Table 53. In this case, there are an infinite number of solutions. If we write x and z in terms of y, the general solution can be represented as

$$(x, y, z) = (1 - y, y, y + 2) = y(-1,1,1) + (1,0,2)$$

which is the equation for a line through the point $(1,0,2)$ and parallel to the vector $(-1,1,1)$.

Table 53. Gaussian Elimination – Infinite Number of Solutions

Equation Format	Row Operations	Matrix Format
$\boldsymbol{x+y=1}$ $\boldsymbol{2x+3y-z=0}$ $\boldsymbol{3x+4y-z=1}$		$\left\langle \begin{matrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ 3 & 4 & -1 \end{matrix} \middle\| \begin{matrix} 1 \\ 0 \\ 1 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y-z=-2}$ $\boldsymbol{y-z=-2}$	$R_2 - 2R_1$ $R_3 - 3R_1$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{matrix} \middle\| \begin{matrix} 1 \\ -2 \\ -2 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y-z=-2}$	$R_3 - R_2$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{matrix} \middle\| \begin{matrix} 1 \\ -2 \\ 0 \end{matrix} \right\rangle$

In some cases, there is no solution. As can be seen in the bottom row of Table 54, we arrive at a contradiction when applying the Gaussian elimination which leads us to conclude that the set of equations is inconsistent and thus has no solution.

Table 54. Gaussian Elimination – No Solution

Equation Format	Row Operations	Matrix Format
$\boldsymbol{x+y=1}$ $\boldsymbol{2x+3y-z=0}$ $\boldsymbol{3x+4y-z=2}$		$\left\langle \begin{matrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ 3 & 4 & -1 \end{matrix} \middle\vert \begin{matrix} 1 \\ 0 \\ 2 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y-z=-2}$ $\boldsymbol{y-z=-1}$	$R_2 - 2R_1$ $R_3 - 3R_1$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{matrix} \middle\vert \begin{matrix} 1 \\ -2 \\ -1 \end{matrix} \right\rangle$
$\boldsymbol{x+y=1}$ $\boldsymbol{y-z=-2}$ $\boldsymbol{0=1}$ (contradiction)	$R_3 - R_2$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{matrix} \middle\vert \begin{matrix} 1 \\ -2 \\ 1 \end{matrix} \right\rangle$

A variation of the Gaussian elimination, known as the Gauss–Jordan elimination, can be used to find the inverse of a matrix. If A is a square matrix, then row reduction can be used to compute the inverse of A. In a starting format similar to the Gaussian elimination, the matrix is placed in the left of the tableaux and the identity matrix is placed to the right. Through a series of row operations, the matrix on the left is transformed to the identity matrix. The same operations are synchronously applied to the matrix on the right. The resulting matrix on the right is the inverse of A (if the inverse exists).

Table 55 shows the Gauss-Jordan elimination as applied to the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ -3 & -3 & 1 \end{bmatrix}$ with the result being $A^{-1} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 3 & 0 & 1 \end{bmatrix}$.

Table 55. Matrix Inversion

Row Operations	Matrix Format
	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ -3 & -3 & 1 \end{matrix} \middle\vert \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right\rangle$
$\boldsymbol{R_2 - 2R_1}$ $\boldsymbol{R_3 + 3R_1}$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{matrix} \middle\vert \begin{matrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{matrix} \right\rangle$
$\boldsymbol{R_2 + R_3}$	$\left\langle \begin{matrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \middle\vert \begin{matrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 3 & 0 & 1 \end{matrix} \right\rangle$
$\boldsymbol{R_1 - R_2}$	$\left\langle \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \middle\vert \begin{matrix} 0 & -1 & -1 \\ 1 & 1 & 1 \\ 3 & 0 & 1 \end{matrix} \right\rangle$

This gives us another way of solving the system of linear equations in Table 52. As shown in Figure 108, we first represent the system of equations in matrix format, and then multiple each side of the equation by A^{-1} to obtain the solution. This works because of the way matrix multiplication is defined.

$$\begin{matrix} x+y=1 \\ 2x+3y-z=0 \\ -3x-3y+z=1 \end{matrix} \Rightarrow \begin{matrix} & A & & & X & & Y \\ \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & -1 \\ -3 & -3 & 1 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} & = & \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \end{matrix} \Rightarrow X = A^{-1}Y = \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix}$$

Figure 108. Solution of System of Linear Equations using Matrix Inversion

15.4 Vector Spaces

15.4.1 Basic Definitions and Theorems

Recall from earlier in this book where Boolean algebra was defined to cover a variety of topics (e.g., propositional logic and set theory) that have similar properties. We are faced with a similar opportunity regarding matrices, i.e., there is a set of matrix properties that can be applied to many other entities. The common structure is known as a vector space and is defined as follows:

A **vector space** V is a non-empty set equipped with an addition operation and a scalar multiplication operation such that for all $\alpha, \beta \in \mathbb{R}$ (i.e., Real numbers) and all $u, v, w \in V$:

- $u + v \in V$ (closure under addition)
- $u + v = v + u$ (the commutative law for addition)
- $u + (v + w) = (u + v) + w$ (the associative law for addition).
- $\exists\, 0 \in V$, called the zero vector, such that $\forall v \in V,\ v + 0 = v$
- $\forall v \in V,\ \exists -v \in V$, called the negative of v, such that $v + (-v) = 0$
- $\alpha v \in V$ (closure under scalar multiplication)
- $\alpha(u + v) = \alpha u + \alpha v$ (distributive law for vectors)
- $(\alpha + \beta)v = \alpha v + \beta v$ (distributive law for scalars)
- $\alpha(\beta v) = (\alpha\beta)v$ (associative law for scalar multiplication).
- $1v = v$ (multiplicative identity)

The following theorem may seem obvious but it does need to be proved (only using the stated properties of vector spaces).

Theorem 15-1 For $v \in V$, where V is a vector space, $0v = 0$ and $(-1)v = -v$.

Proof: We have that $0v = (0 + 0)v = 0v + 0v$ by the definition of the zero vector and the distributive law for scalars. Thus, $0v = 0v + 0v$. Now add $-0v$ to both sides of the previous equation to get

$0 = 0v + (-0v) = (0v + 0v) + (-0v)$	by definition of 0, $0 = 0v + (-0v)$
$= 0v + \big(0v + (-0v)\big)$	by the associative law
$= 0v + 0$	using the definition of negative for a vector
$= 0v$	by definition of zero vector

Thus, $0 = 0v$.

Now for the second part of the theorem, we make use of the fact that $0 = 1 + (-1)$ in the following:

$$0 = 0v = \big(1 + (-1)\big)v = 1v + (-1)v = v + (-1)v$$

By the definition of the negative of a vector, $-v = (-1)v$. So, if you multiply a vector v by the scalar -1, you get the negative of v■

A **subspace** W of a vector space V is a non-empty subset of V that is also a vector space under the same operations of addition and scalar multiplication as defined for V.

Theorem 15-2 A non-empty subset W of a vector space V is a subspace if and only if W is closed under addition and scalar multiplication, i.e.,

- if $w_1, w_2 \in W$ then $w_1 + w_2 \in W$
- if $\alpha \in \mathbb{R}$ and $w \in W$ then $\alpha w \in W$.

Proof: If W is a subspace of V, then by the definition of subspace, W is a vector space in its own right, and thus the two closure properties hold.

Going in the other direction, assume W is a subset of V and the two closure properties hold for W. Most of the vector space properties of V directly carryover to W. However, a few of the properties require further explanation:

- Since W is non-empty, there exist $w \in W$. By Theorem 15-1, $0w = 0$ and by the scalar closure property on W, $0 \in W$.
- By Theorem 15-1 and the scalar closure property on W, we have for any $w \in W$, $(-1)w = -w \in W$■

15.4.2 Examples of Vector Spaces

15.4.2.1 Matrices

The set of $m \times n$ matrices whose entries are real numbers is a vector space, with addition and scalar multiplication of matrices as defined in Section 15.1. The zero vector is the $m \times n$ matrix with all entries equal to 0. The negative of a matrix A is -1 times A.

As a byproduct of $m\ x\ n$ matrices being a vector space, we also have that n-dimensional real space (denoted $\mathbb{R}^n$) is a vector space. The elements of $\mathbb{R}^n$ can be viewed as either $1\ x\ n$ or $n\ x\ 1$ matrices.

15.4.2.2 Functions

Consider the set F of functions from $\mathbb{R}$ to $\mathbb{R}$ where addition and multiplication are defined as follows:

- For $f, g \in F$, $f + g$ is defined to be $(f + g)(x) = f(x) + g(x)$.
- For $f \in F$ and scalar $\alpha \in \mathbb{R}$, αf is defined to be $(\alpha f)(x) = \alpha f(x)$.

The zero vector in F is the function $z(x) = 0, \forall x \in \mathbb{R}$. With the above operations, F is a vector space.

15.4.2.3 Infinite Sequences of Real Numbers

The set S of all infinite sequences of real numbers, $x = \{x_1, x_2, x_3, \dots\}$, is a vector space. The notation $x = \{x_n\}, n \geq 1$ is used as a shorthand for such a sequence. For example, the sequence $x = \{1, 3, 5, 7, \dots\}$ can also be represented as $\{x_n\}$ with $x_n = 2n + 1, n = 0,1,2,3, \dots$ Addition and scalar multiplication are defined as follows:

- For $x = \{x_n\},\ y = \{y_n\} \in S$, then $x + y$ is defined as $\{x_n + y_n\}$.
- For $x = \{x_n\}$ and $\alpha \in \mathbb{R}$, the scalar multiplication is defined as $\alpha x = \{\alpha x_n\}$.

15.5 Linear Independence, Linear Transformations and Bases

For vectors $v_1, v_2, \dots, v_n$ in a vector space V, the vector $v = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$ is a **linear combination** of the said vectors. The scalars a_i are called coefficients.

We've already seen the concept of a linear combination regarding real numbers (basically, a one-dimensional vector space) in the number theory section of this book, and in the context of functions (which, as noted in the previous section, can also be viewed as a vector space).

In Section 15.3.2, we saw how a system of equations could be represented in matrix format and then solved via the Gaussian elimination. There is a third way to represent a system of equations, i.e., as a linear combination of vectors (as shown at the bottom of Figure 109).

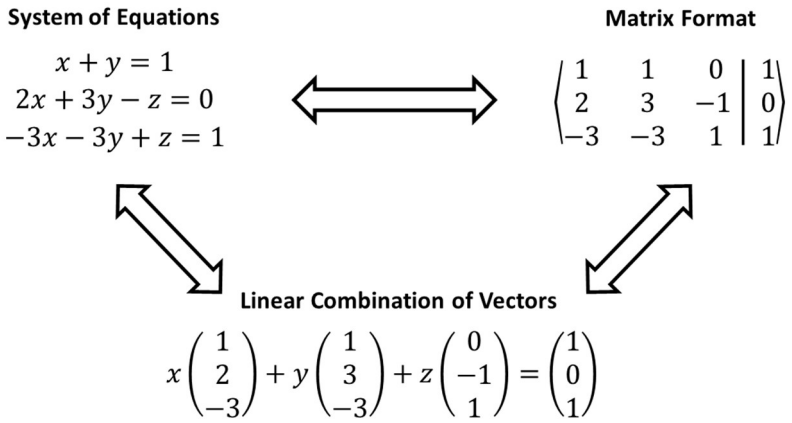


Figure 109. Multiple Ways of Expressing a System of Equations

Let $v_1, v_2, \dots, v_k \in V$ where V is a vector space. The vectors $v_1, v_2, \dots, v_k$ are said to be **linearly independent** if and only if the vector equation $a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0$ has the unique solution $a_1 = 0, a_2 = 0, \dots, a_k = 0$. If solutions other than all the coefficients equal to zero exist, then the set of vectors are said to be **linearly dependent**.

The three vectors shown at the bottom of Figure 109 (label them as v_1, v_2 and v_3) are linearly independent. This can be shown by determining that the only solution to $xv_1 + yv_2 + zv_3 = 0$ is the $x = y = z = 0$. Using the matrix format equivalent, Table 56 shows that the unique solution is the zero vector.

Table 56. Linear Independence Example

Row Operations	Matrix Format
	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 \\ -3 & -3 & 1 & 0 \end{array} \right\rangle$
$\boldsymbol{R_2 - 2R_1}$ $\boldsymbol{R_3 + 3R_1}$	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right\rangle$
$\boldsymbol{R_2 + R_3}$	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right\rangle$
$\boldsymbol{R_1 - R_2}$	$\left\langle \begin{array}{ccc\|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right\rangle$

As an example of linear dependence, consider the vectors $v_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, v_2 = \begin{bmatrix}1\\3\\4\end{bmatrix}$ and $v_3 = \begin{bmatrix}0\\-1\\-1\end{bmatrix}$.

In Table 57, the equation $xv_1 + yv_2 + zv_3 = 0$ is put into matrix format and then solved. The solution can be read off as

$$x + y = 0$$
$$y - z = 0$$

The value of variable y is free to choose as we wish. The above can be rewritten in vector form as

$$\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}-y\\y\\y\end{bmatrix} = y\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

which is a straight line in $\mathbb{R}^3$. Thus, we have an infinite number of non-zero solutions to the equation $xv_1 + yv_2 + zv_3 = 0$, and the set of vectors v_1, v_2 and v_3 is linear dependent.

Table 57. Linear Dependence Example

Row Operations	Matrix Format
	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 \\ 3 & 4 & -1 & 0 \end{array} \right\rangle$
$\boldsymbol{R_2 - 2R_1}$ $\boldsymbol{R_3 - 3R_1}$	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{array} \right\rangle$
$\boldsymbol{R_3 - R_2}$	$\left\langle \begin{array}{ccc\|c} 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right\rangle$

Theorem 15-3 A set of vectors $v_1, v_2, \ldots, v_n$ in a vector space V is linearly dependent if and only if at least one vector in the set is a linear combination of the others.

Proof: By definition, the set $v_1, v_2, \ldots, v_n$ is linear dependent if there exists a non-zero solution to $a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0$, i.e., at least one of the coefficients does not equal zero (say $a_i \neq 0$). Thus, we can write

$$v_i = \frac{a_1}{a_i}v_1 + \frac{a_2}{a_i}v_2 + \ldots + \frac{a_{i-1}}{a_i}v_{i-1} + \frac{a_{i+1}}{a_i}v_{i+1} + \cdots + \frac{a_n}{a_i}v_n$$

So, v_i can be written as a linear combination of the other vectors in the set.

Going in the other direction, assume one of the vectors (say v_i) can be written as a linear combination of the other vectors, i.e.,

$$v_i = b_1v_1 + b_2v_2 + \ldots + b_{i-1}v_{i-1} + b_{i+1}v_{i+1} + \cdots + b_nv_n$$

which can be rearranged as

$$b_1v_1 + b_2v_2 + \ldots + b_{i-1}v_{i-1} - v_i + b_{i+1}v_{i+1} + \cdots + b_nv_n = 0$$

Thus, there exists a non-zero solution to the equation $a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0$, i.e., the set $v_1, v_2, \ldots, v_n$ is linear dependent■

In vector spaces, linear combinations are unique. In other words, if a vector v can be expressed as a linear combination of a set of linearly independent vectors, this can only be done in one way. More formally, we state the result in the following theorem.

Theorem 15-4 If $v_1, v_2, \ldots, v_n$ is a set of linearly independent vectors in a vector space V and if

$$a_1v_1 + a_2v_2 + \ldots + a_nv_n = b_1v_1 + b_2v_2 + \ldots + b_nv_n$$

then $a_i = b_i, for\ i = 1,2, \ldots, n.$

Proof: The equation

$$a_1v_1 + a_2v_2 + \ldots + a_nv_n = b_1v_1 + b_2v_2 + \ldots + b_nv_n$$

implies

$$(a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \ldots + (a_n - b_n)v_n = 0$$

However, we have been given that the set of vectors $v_1, v_2, \ldots, v_n$ is linearly independent, and so all the coefficients in the above equation must be equal to zero which implies $a_i = b_i$, for $i = 1,2, \ldots, n$■

Related to the concept of linear independence is that of a basis for a vector space. The general idea is to determine a minimal set of vectors for which all other vectors in a given vector space can be represented as a linear combination.

A set of vectors S **spans** a vector space V, if any vector $v \in V$ can be expressed as a linear combination of the vectors in S. If, in addition, the set of vectors is linearly independent, then the set of vectors is said to be a **basis** for V. The number of elements in a basis for a vector space is referred to as the **dimension** of the vector space. If the basis has an infinite number of elements, the vector space is said to be infinite dimensional. Some examples of infinite dimensional vector spaces:

- The set of polynomials in one variable
- The set of continuous functions from $\mathbb{R}$ to itself.
- The set of all differentiable functions from $\mathbb{R}$ to itself. More generally, any collection of functions which are closed under addition and scalar multiplication.
- The set of all the infinite sequences over $\mathbb{R}$.

Theorem 15-5 If a set of vectors B in a finite dimensional vector space V is a basis, then every element in V can be written uniquely as a linear combination of vectors in B.

Proof: This follows immediately from the definition of a basis and Theorem 15-4■

The following theorem is stated without proof.

Theorem 15-6 Let V be a vector space with basis $B = \{v_1, v_2, \dots, v_n\}$ then

- *Any subset of V containing more than n vectors must be linearly dependent.*
- *Any subset of V containing less than n vectors cannot span V.*

A consequence of Theorem 15-6 is that all bases of a finite dimensional vector space have the same number of elements.

For example, consider the vector space $\mathbb{R}^n$ consisting of vectors of the form $(v_1, v_2, \dots, v_n)$ where $v_i \in \mathbb{R}$ for $i = 1,2, \dots, n$. The set of vectors $e_1 = (1,0, \dots,0), e_2 = (0,1, \dots 0), \dots, e_n = (0,0, \dots,1)$ is a basis for $\mathbb{R}^n$ and is referred to as the **standard basis** for $\mathbb{R}^n$.

- Any vector $(v_1, v_2, \dots, v_n)$ can be written as $v_1\boldsymbol{e_1} + v_2\boldsymbol{e_2} + \dots + v_n\boldsymbol{e_n}$ and so, $\{e_1, e_2, \dots, e_n\}$ spans $\mathbb{R}^n$.
- If one forms the matrix equivalent of the vector equation $x_1\boldsymbol{e_1} + x_2\boldsymbol{e_2} + \dots + x_n\boldsymbol{e_n} = 0$, it is apparent that the only solution is $(x_1, x_2, \dots, x_n) = (0,0, \dots,0)$. Thus, $\{e_1, e_2, \dots, e_n\}$ is linearly independent.

The above bullet items prove that $\{e_1, e_2, \dots, e_n\}$ is a basis for $\mathbb{R}^n$.

A function from one vector space V to another vector space W is a mapping which assigns to every vector $v \in V$ a unique vector $w \in W$. If the function is linear, then it is known as a linear transformation. More formally, given vector spaces V and W, a function $L: V \longrightarrow W$ is a **linear transformation** if $\forall u, v \in V$ and $\forall \alpha \in \mathbb{R}$:

- $L(u + v) = L(u) + L(v)$
- $L(\alpha u) = \alpha L(u)$

As an example, consider any $m\ x\ n$ matrix A whose elements are in $\mathbb{R}$. Define the function L as $L(v) = Av$ where v is an $n\ x\ 1$ matrix whose elements are in $\mathbb{R}$. The dimension of Av is $m\ x\ 1$. L is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ since

- $L(u + v) = A(u + v) = Au + Av = L(u) + L(v)$
- $L(\alpha u) = A(\alpha u) = \alpha Au = \alpha L(u)$.

If U and V are vector spaces and there exist a bijective linear transformation $L: V \longrightarrow U$, then L is called an isomorphism, and U and V are said to be **isomorphic**. (Recall that bijective functions were defined in Section 8.1.)

Theorem 15-7 If U and V are vector spaces of the same finite dimension n, then U and V are isomorphic.

Proof: Let $A = \{u_1, u_2, \dots, u_n\}$ be a basis for U and $B = \{v_1, v_2, \dots, v_n\}$ be a basis for V. We know from Theorem 15-5 that every vector $u \in U$ can be written uniquely as a linear combination of the vectors in A, i.e., $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$. Next, define a mapping L from U to V as follows:

- $L(u_i) = v_i$
- $\forall u \in U, L(u) = a_1v_1 + a_2v_2 + \cdots + a_nv_n$ where $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$ is a linear combination of the basis A.

We want to show that L is a bijective linear transformation from V to U.

First, note that L is a function from V to U since it maps each vector in V to exactly one vector in U.

For $u, w \in U$ where $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$ and $w = b_1u_1 + b_2u_2 + \cdots + b_nu_n$, we have

$$L(u + w) = L\big((a_1 + b_1)u_1 + (a_2 + b_2)u_2 + \cdots + (a_n + b_n)u_n\big)$$
$$= (a_1 + b_1)v_1 + (a_2 + b_2)v_2 + \cdots + (a_n + b_n)v_n$$
$$= (a_1v_1 + a_2v_2 + \cdots + a_nv_n) + (b_1v_1 + b_2v_2 + \cdots + b_nv_n)$$
$$= L(u) + L(w)$$

For $\alpha \in \mathbb{R}$ and $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$, we have

$$L(\alpha u) = L(\alpha a_1u_1 + \alpha a_2u_2 + \cdots + \alpha a_nu_n)$$
$$= \alpha a_1L(u_1) + \alpha a_2L(u_2) + \cdots + \alpha a_nL(u_n)$$
$$= \alpha a_1v_1 + \alpha a_2v_2 + \cdots + \alpha a_nv_n$$
$$= \alpha L(u)$$

Thus, L is a linear transformation.

Next, take any vector $v = a_1v_1 + a_2v_2 + \cdots + a_nv_n \in V$. The vector $u = a_1v_1 + a_2v_2 + \cdots + a_nv_n \in U$ is mapped to v by L. Thus, L is surjective.

Finally, we need to show L is injective. Let $L(u) = L(w)$, where $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$ and $w = b_1u_1 + b_2u_2 + \cdots + b_nu_n$. Then $L(u) = a_1v_1 + a_2v_2 + \cdots + a_nv_n = b_1v_1 + b_2v_2 + \cdots + b_nv_n = L(w)$ which implies $(a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_n - b_n)v_n = 0$. However, $\{v_1, v_2, \ldots, v_n\}$ is a basis for V, and thus linearly independent, which implies $a_i = b_i$ for $i = 1,2, \ldots n$. So, $u = w$ and we have shown that L is injective.

In summary, L is a bijective linear transformation from U to V and thus, by definition, U and V are isomorphic■

15.6 Exercises

1. Referring to Figure 101, find C^t and then multiple it times A, i.e., find C^tA.
2. Write the line $y = -2x + 5$ in terms of a vector sum (similar to the approach taken in Figure 106).
3. Use the Gaussian elimination to solve the following system of linear equations:

$$x + 3y + z = 5$$
$$2x + y - z = 2$$
$$3x - 5y + 2z = 0$$

4. In the example related to Table 55, verify via matrix multiplication that A^{-1} is in fact the inverse of A.

5. Find the inverse of $\begin{bmatrix}1 & 2 & 1\\ 1 & 3 & 2\\ 2 & 2 & 1\end{bmatrix}$.
6. Show that set S of all infinite sequences of real numbers (as defined in Section 15.4.2) is a vector space.
7. Determine whether the vectors $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}$ and $\begin{bmatrix}7\\8\\9\end{bmatrix}$ are linearly independent or not.
8. Show that the set of vectors $\{(1,2),(2,5)\}$ form as basis for $\mathbb{R}^2$. ***Hint****: Show the two vectors are linearly independent and then use Theorem 15-6 along with the fact that* $\mathbb{R}^2$ *is of dimension 2.*
9. Verify that mapping L defined by $L\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right) = \begin{bmatrix}a\\b\\c\\d\end{bmatrix}$ is a linear transform from the vector space of 2 x2 matrices with real-valued elements to the vector space of 4 x 1 column vectors.

16 Proofs

This section provides a summary and consolidation of the ideas related to the various types of proofs presented earlier in the book. For the interested reader, there are several books devoted entirely to techniques of mathematical proofs, i.e.,

- George Polya's book "How to Solve It" [72] (reprint of the original classic from 1945)
- Andrew Wohlgemuth's book "Introduction to Proof in Abstract Mathematics" [73]
- Ted Sundstrom's book "Mathematical Reasoning: Writing and Proof" [74].

16.1 Direct Proof

In a direct proof, the conclusion is deduced by making use of existing axioms and definitions, and previously proven theorems.

For example, Theorem 4-1 (concerning basic laws of propositional logic) entails several very basic direct proofs.

Some other examples of direct proofs:

- Theorem 5-1
- Theorem 6-1 provides both a direct and indirect proof
- Theorem 6-12
- Theorem 14-9.

16.2 Mathematical Induction

As noted earlier, mathematical induction is not to be confused with inductive reasoning, but is rather an example of deductive reasoning. This type of proof is common in number theory, combinatorics and many other branches of mathematics. Recall the general approach, i.e., prove the theorem holds for an initial case (e.g., $n = 1$) and then show if the theorem holds for $n = k$, it must hold for the case $n = k + 1$.

Some examples of proof by mathematical induction:

- Theorem 9-4 (binomial theorem)
- Theorem 9-17
- Theorem 12-6
- Theorem 12-10
- Theorem 14-10.

Mathematical induction is considered to be a type of direct proof.

16.3 Constructive Proof

As the name suggests, in a constructive proof, the entity being claimed to exist is actually exhibited (constructed).

Theorem 14-6, concerning the existence of an r-regular graph of order n, is a constructive proof since the proof tells the reader how to construct such a graph.

Euclid's algorithm (see Theorem 9-12), concerning the determination of a greatest common divisor for two integers, is another example of a constructive proof. In general, theorems related to algorithms are constructive in nature.

Kruskal's algorithm (Section 14.4), concerning the creation of a minimal spanning tree of a graph, is another example of a constructive proof. In this case, we did not state a theorem but rather, just presented the algorithm.

In a negative sense, the creation of a counterexample to disprove a proposition can also be classified as constructive. For example, the cantor diagonalization argument (Section 6.6.2) involves the creation of a counterexample to prove there exists a set which is not countable.

Constructive proofs are also considered to be direct proofs.

16.4 Existence Proof

Existence proofs demonstrate that a given entity must exist but does not necessarily tell one how to construct the entity. It is fair to say that all constructive proofs are also existence proofs, but not the other way around.

As an example of an existence proof that is not constructive, consider the statement "the function $f(x) = x^3 + 2$ has a real root, i.e., there is a solution to $f(x) = 0$ that is a real number." The proof goes as follows:

> As x goes to infinity so does $f(x)$, and as x goes to minus infinity, so does $f(x)$. Further, since $f(x)$ is continuous and its values range from minus infinity to plus infinity, it must take on every real value at least once (including 0).

So, we have established that $f(x) = 0$ must have a real-valued solution, but have not determined the specific value of x that is the solution.

16.5 Proof by contraposition

Consider a proposition to be proved of the form "If A, then B" or more concisely, $A \Rightarrow B$. As was noted in Section 4.9.4, $A \Rightarrow B$ is equivalent to $\neg B \Rightarrow \neg A$ (see Table 7 for the truth table that proves this fact). This was referred to as the rule of contraposition. In some cases, it is easier to prove the contrapositive form of a proposition.

For example, consider the following proposition: For a positive integer x, if x^3 is odd, then x is odd.

> **Proof**: Assume x is even, i.e., of the form $x = 2n$ for some integer n, then $x^3 = 8n^2$ which is divisible by 2 and thus, even. So, if x is not odd (i.e., even), then x^3 is also not odd (i.e., even)■

There are other equivalents to $A \Rightarrow B$, e.g., the rule of material implication states that the statement $A \Rightarrow B$ is equivalent to the statement $\neg A \vee B$. Thus, proving one of the two statements implies that the other statement is also true.

16.6 Proof by contradiction (Reductio ad absurdum)

From the Wikipedia article entitled "Reductio ad absurdum" [75]:

> In logic, *reductio ad absurdum* (Latin for '"reduction to absurdity"'), also known as *argumentum ad absurdum* (Latin for "argument to absurdity"), apagogical arguments, negation introduction or the appeal to extremes, is a form of argument that attempts to establish a claim by showing that the opposite scenario would lead to absurdity or contradiction. It can be used to disprove a statement by showing that it would inevitably lead to a ridiculous, absurd, or impractical conclusion, or to prove a statement by showing that if it were false, then the result would be absurd or impossible.

Assume the goal is to prove the statement $A \Rightarrow B$ is true. In proof by contradiction, statement A is assumed to be true and B assumed to be false (i.e., $\neg B$ to be true). Next, a contraction (i.e., a falsehood) is derived which effectively proves the statement $\neg B \Rightarrow False$ to be a true statement. However, the statement $\neg B \Rightarrow False$ is only true when $\neg B$ is also false (see the truth table for $\Rightarrow$, i.e., Table 6 on Page 27). So, B must be true.

This is a different approach from proof by contraposition where $\neg B$ is assumed true and one proves that $\neg A$ is true.

An often cited example of proof by contraction is the irrationality of $\sqrt{s}$ when s is a prime number (see Theorem 9-19). Theorem 9-21 is another classic example of proof by contradiction.

Proof by contradiction and proof by contraposition are sometimes classified as indirect proofs.

16.7 Decomposition and Levels

More complex proofs may involve several of the above methods of proof. In general, a basic principle in mathematics is to divide a problem into subproblems (preferably ones that already have solutions). The subproblems could very well be solved (proved) using different approaches.

To take an extreme example, consider Fermat's Last Theorem which states that equations of the form $x^n + y^n = z^n$ do not have positive integer solutions (in x, y, z) for $n \geq 3$ where n is an integer. The problem is attributed to Pierre de Fermat who first posed the problem in 1637. The conjecture remained unproven for 358 years until Andrew Wiles provided a proof in 1994 (published in 1995 [76]). The proof from Wiles goes on for 109 pages and involves the proof of many subproblems and extensive references to already existing (and proven) theorems. The overall strategy was proof by contradiction. The Wikipedia article "Wiles's proof of Fermat's Last Theorem" [77] describes the high-level parts of the proof.

16.8 Theorem, Lemma and Corollary

While the terms "lemma" and "corollary" are not used much in this book, the terms do often appear in many mathematics books that involve proofs. The following definitions are typical of how mathematicians label results:

> **Theorem** – a mathematical result that is proved using logical reasoning. In mathematical articles and books, the term theorem is often reserved for the most important results.
>
> **Lemma** – a result whose sole purpose is to help in proving a theorem. A lemma is typically a stepping-stone on the way to proving a theorem. There are, however, many cases where a result known as a lemma takes on a life of its own (see the extensive list of famous lemmas in the Wikipedia article [78]).

Corollary – a result in which the (usually short) proof relies primarily on a given theorem.

To be clear, lemmas and corollaries are also proved using logical reasoning.

17 Algorithms

17.1 Overview

The Definitive Glossary of Higher Mathematical Jargon provides the following definition of "**algorithm**" [79]

> A finite series of well-defined, computer-implementable instructions to solve a specific set of computable problems. It takes a finite amount of initial input(s), processes them unambiguously at each operation, before returning its outputs within a finite amount of time.

In other words, an algorithm is basically a set of steps for solving a given problem. The above definition adds the criterion "computer-implementable" which is a bit of a moving target as we create more advanced Artificial Intelligence (AI) based entities.

We've already seen many examples of algorithms in this book, e.g., Euclid's algorithm for computing the gcd of two numbers, the division algorithm (Theorem 9-6), Kruskal's algorithm for determining a minimum spanning tree in a graph, the Gaussian elimination for solving a set of linear equations and the Gauss–Jordan elimination for finding the inverse of a matrix.

[Author's Remark: "What set of instructions is not an algorithm?" If we modify the above definition to say "instructions which in theory can be automated", then perhaps all instruction sets can be considered to be algorithms. This would include things that are now not possible with current AI. I hesitate to give an example as it will likely be wrong in a few years if not sooner.

If we include humans as part (or all) of the automation, then items such as Do It Yourself (DIY) instructions, long-division (by hand) and recipes for food would also be examples of algorithms.

Then there are complex tasks, e.g., raising a child or deciphering a newly discovered ancient text written in an unknown language. These are learned activities that entail a lot of improvisation. Even complex tasks of this nature may be "automated" but not in the sense of a predetermined set of instructions but rather by creating sentient beings that can learn and adapt.**]**

17.2 Classification

17.2.1 Recursive and Iterative Algorithms

Recursive and iterative algorithms both entail the repeated execution of a set of instructions.

- A recursive algorithm is one that invokes other instances of itself until certain conditions are met, regarding the solution of a given problem.
- An iterative algorithm executes a loop repeatedly until the controlling condition becomes false, e.g., until some counter n is greater than 100 or some other positive integer.

In general, recursive algorithms are easier to write, but they do not perform well as compared to iterative algorithms. On the other hand, iterative algorithms are relatively harder to write than recursive algorithms.

Some problems are best solved by one approach over another. For example, the Towers of Hanoi problem (discussed below) is well suited for a recursive solution. Every recursive algorithm has an iterative equivalent, and vice versa.

17.2.1.1 Fibonacci Sequence

The Fibonacci sequence provides a simple example of a recursive algorithm. The first two values of the sequence are given by $Fib(1) = 1$ and $Fib(2) = 1$. The n^{th} element in the sequence is given by the function $Fib(n) = Fib(n-1) + Fib(n-2)$. For example, to determine $Fib(5)$ recursively, do the following:

- $Fib(5) = Fib(4) + Fib(3)$ but we don't yet know $Fib(4)$ or $Fib(3)$ and so, need to recursively call the Fib function (algorithm) for these two cases.
    - $Fib(3) = Fib(2) + Fib(1)$. We know $Fib(1)$ but need to call the Fib function again to determine $Fib(2) = Fib(1) + Fib(0) = 1 + 1 = 2$ and so $Fib(3) = 2 + 1 = 3$
    - $Fib(4) = Fib(3) + Fib(2)$ and we recursively, call the Fib function several times to get $Fib(4) = 3 + 2 = 5$.
- We "pop-up" (return) to the initial function call and get $Fib(5) = 5 + 3 = 8$.

The following Python code implements a recursive algorithm for computing the n^{th} Fibonacci number:

```
# Recursive algorithm to find nth Fibonacci number
def Fibonacci(pos):
    #check for the terminating condition
    if pos == 1 :
        #Return the first Fibonacci number, i.e., 1
        return 1
    if pos == 2:
        #Return the second Fibonacci number, i.e., 1
        return 1
    # Calculate the (n-1)th number by calling the function itself
    n_1 = Fibonacci( pos - 1 )
    # Calculation  the (n-2)th number by calling the function itself again
    n_2 = Fibonacci( pos - 2 )
    # Calculate the requested Fibonacci number
    n = n_1 + n_2
    # Return the requested Fibonacci number
    return n
#For example, calculate the 5th Fibonacci
nth_fib = Fibonacci(5)
print (nth_fib)
```

At first look, the above code may seem like magic. The key is to understand the two lines shown in red. These function calls invoke yet other function calls (recursively) until they reach the two given Fibonacci numbers in the sequence, i.e., 1,1. Figure 110 shows the function calls involved in calculating the 5th Fibonacci number. The function calls go all the way to the leaves of the tree before returning in the reverse direction. Note the inefficiency in calling Fibonacci(2) three separate

times. The algorithm is quite slow. Try replacing “5” with “33” (takes a few seconds on a home computer) and then try “43” (takes several minutes on a home computer).

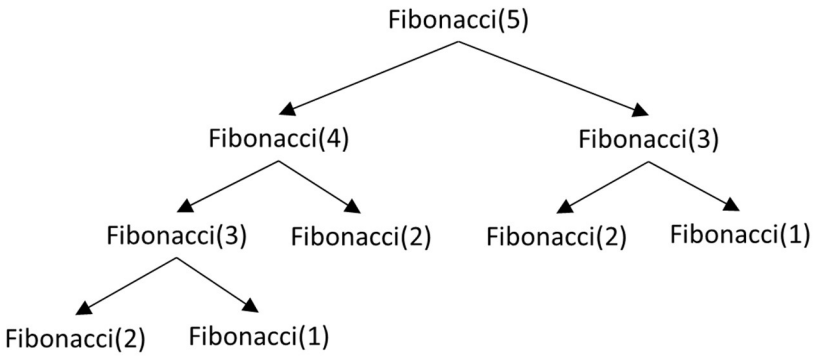


Figure 110. Recursive Function Calls for Fibonacci Sequence

It is also possible to write an iterative algorithm for computation of the n^{th} Fibonacci number. The following code in Python does just that.

```
# Iterative algorithm to find nth Fibonacci number
def Fibonacci(n):
    # The first two Fibonacci numbers.
    fib_prev_2 = 1
    fib_prev_1 = 1
    for i in range(1,n-1):
        fib = fib_prev_1 + fib_prev_2
        fib_prev_2 = fib_prev_1
        fib_prev_1 = fib
    #Return the requested Fibonacci number
    return fib
#For example, calculate the 7th Fibonacci
nth_fib = Fibonacci(7)
print (nth_fib)
```

In the above code, fib_prev_1 and fib_prev_2 are used to store the previous 2 Fibonacci numbers that have been calculated. The next Fibonacci number is then calculated, i.e., fib = fib_prev_1 + fib_prev_2. Next, we reassign fib_prev_1 to fib_prev_2, assign fib to fib_prev_1 and then calculate the next Fibonacci number. This “for loop” is repeated until the desired Fibonacci number has been calculated. Now, replace “7” with “43”. The answer will be almost immediate.

17.2.1.2 Towers of Hanoi

The Towers of Hanoi is a classic problem that shows the power of recursion. The problem goes like this:

- There are n disks (each of a different diameter) on a peg, with the largest disk on the bottom and no larger disk above a smaller disk, i.e., the disks are stacked in descending order of their diameters.

- There are three pegs.
- The n disks on the peg are to be moved to one of the other two pegs.
- Only one disk at a time can be moved.
- It is not allowed to put a large disk on top of a smaller disk.

The starting position for the problem is shown in Figure 111.

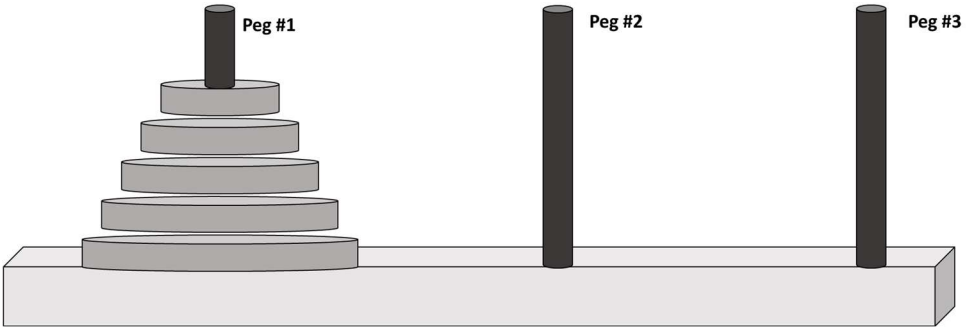


Figure 111. Towers of Hanoi

The basic idea is to reduce the problem to several simpler problems. For the example in the figure, we don't know how to move 5 disks from Peg #1 to #3, but if we could call a function to move the top four disks from Peg #1 to #2, move the remaining disk from Peg #1 to #3 and then call a function to move the 4 disks from Peg #2 to #3, we would be done. So, we reduced the problem from moving 5 disks to two simpler problems of moving 4 disks. The 4-disk problem can similarly be reduced to 3-disks problems and so on. The following Python code (taken from the Wikipedia article on the Tower of Hanoi [80]) implements a recursive algorithm for solving the Towers of Hanoi problem for 5 disks. The code can be adapter to n disks by extending the array A size n, and changing the last line to $move(n, A, C, B)$.

```
A = [5,4,3,2,1]
B = []
C = []
def move(n, source, target, auxiliary):
   if n > 0:
      # Move n - 1 disks from source to auxiliary, so they are out of the way
      move(n - 1, source, auxiliary, target)
      # Move the nth disk from source to target
      target.append(source.pop())
      # Display our progress
      print(A, B, C, '_________________', sep='\n')
      print(' ')
      # Move the n - 1 disks that we left on auxiliary onto target
      move(n - 1, auxiliary, target, source)
# Initiate call from source A to target C with auxiliary B
move(5, A, C, B)
```

The previously referenced Wikipedia article also provides an iterative solution to the problem.

17.2.2 Serial and Parallel Algorithms

In most textbooks, algorithms are typically discussed with an assumption that computers execute one instruction at a time, i.e., a serial algorithm. Parallel algorithms, on the other hand, work on the assumption that multiple instructions can be executed independently and at the same time, with the final result being constructed at the end of the process. Parallel algorithms can run on computers with multiple processors, or they can run on multiple separate computers.

The follow are examples of where parallel algorithms may be appropriate versus serial algorithms:

- Given the large size of the graphs for some minimum spanning tree problems, parallel algorithms have been developed, see the Wikipedia article “Parallel algorithms for minimum spanning trees” [81].
- Distributed computing (sometimes referred to as “grid computing”) is a type of parallel computing that makes use of donated time from separate computers connected to a network. For example, individuals may donate some of the idle time on their personal computer to a project. For additional details and a list of many examples, see the Wikipedia article on “Grid computing” [82].
- Parallel sorting entails the arranging of a large number of items based on some predefined order (e.g., alphabetical order). Extensive coverage of many different parallel sorting algorithms can be found in the book “Parallel Sorting Algorithms” [83].
- Distributed web crawling is a computing technique whereby Internet search engines use many computers to index the Internet via “web crawling.” This is essentially a parallel algorithm for indexing the web.
- Many cryptographic protocols are based on the difficulty of factoring large composite integers, often with large prime factors. One approach is two divide the factorization problem into many subproblems and use a parallel algorithm to solve each of the subproblems. For example, something called the “general number field sieve” makes use of distributed computing to factor very large integers, see the Wikipedia article “General number field sieve” [84].

17.3 Graphical Representation

Before one implements an algorithm (talking about computer-based implementations here), it is a good idea to sketch an outline of the steps. This is particularly important (indispensable really) if the algorithm is complex with many steps and needs to be divided into parts (subroutines or similar) for implementation.

17.3.1 Flow Charts

Flow charts are perhaps the best known and most widely used graphical notation for the representation of algorithms. A flowchart shows the steps of an algorithm via interconnected boxes of various types. The flowchart “language” has been an international standard for many years, see ISO 5807:1985 [85].

A subset of the flow chart graphical symbols is shown in Figure 112.

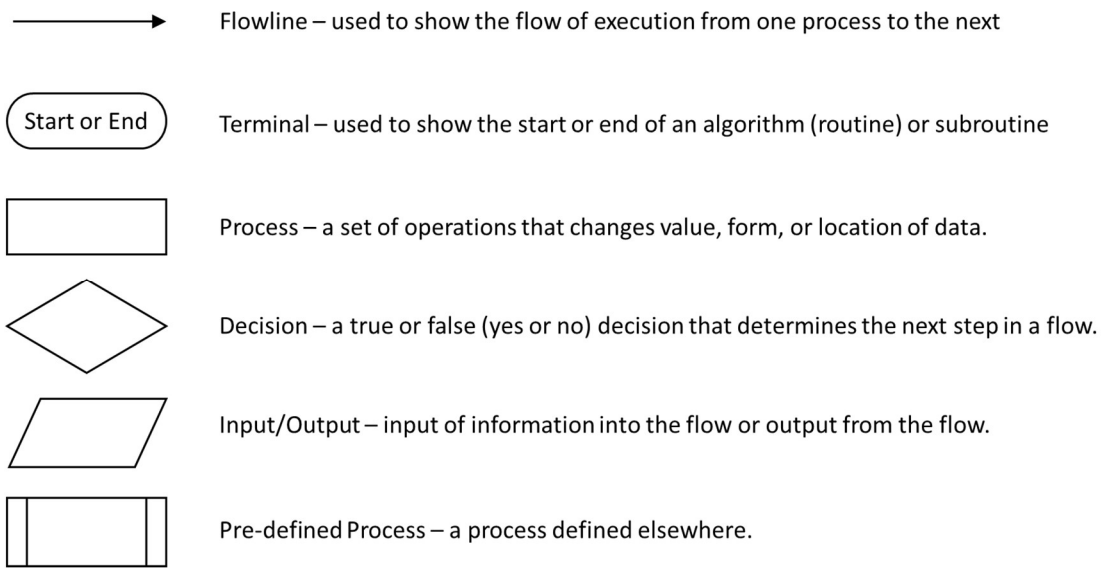


Figure 112. Flow Chart Symbols

The diagram on the left of Figure 113 depicts a flow chart for Kruskal's algorithm. The step concerning determination of the next edge to add to the spanning tree S is expanded in the flow chart on the right of the figure.

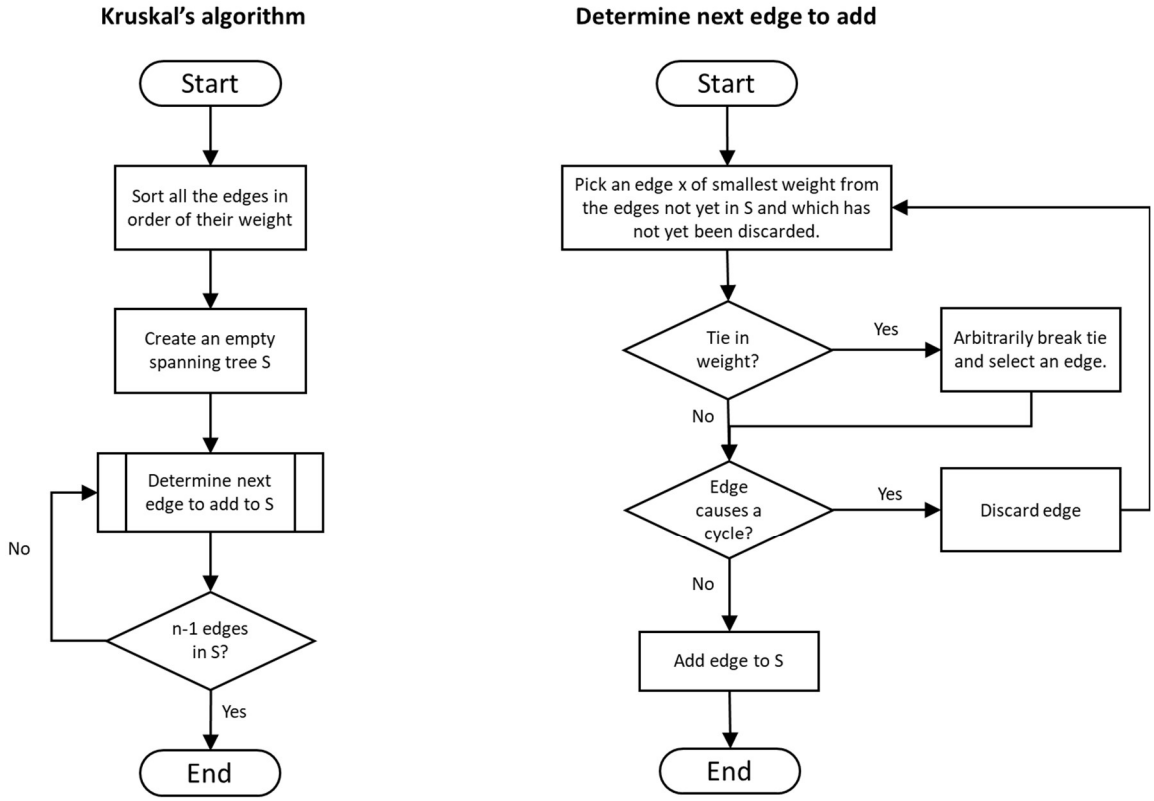


Figure 113. Flow Chart for Kruskal's Algorithm

17.3.2 Other Design Approaches for Algorithms

Structograms (or Nassi–Shneiderman diagrams [86]) support a top-down design approach, where a given problem is reduced into smaller and smaller subproblems, until only simple statements and control flow constructs remain.

DRAKON [87] is a graphical programming and modeling language developed within the Buran space project in Russia. The language provides a way to represent process flows of varying complexity that are easy to read and understand. The graphical aspect of DRAKON is similar to flow charts in concept but different graphic symbols are used.

18 Universal Laws of Mathematics

18.1 Overview

This section covers several mathematical laws that hold true under a variety of conditions. Some of the laws have actually been proven formally under given conditions (e.g., the law of large numbers) while other laws seem to hold true but there is yet no proof.

18.2 Law of Large Numbers

The law of large numbers was introduced earlier in this book, i.e., in Section 12 concerning probability. Recall that this law states that as the number of samples from a given population increases, the average of the samples for a given characteristic (parameter) approaches the expected value of the given population for the characteristic.

Insurance companies benefit from and make use of the law of large numbers. For example, an insurer of automobile drivers sets their rates based on the expectation that the drivers they insure (population sample) will incur an average cost (from liability, collision, medical treatment related to accidents) that is close to the known expected value of such costs for the entire population of drivers. Further, it is typical to focus on subsets of the driver population, e.g., male drivers between 18 and 25 years old, since the expected value of costs will vary greatly for various segments of the driver population. The price of the insurance policy is set accordingly (allowing for some profit).

Another example is diversification of one's holding in the stock market. In order to limit risk, an individual may want to purchase a small amount of each stock in a given category (e.g., value) or index (e.g., Standard and Poor's 500) via a mutual fund. The mutual fund is basically a sample from a large population and thus, by the law of large numbers is expected to have an average gain (or loss) that approaches the entire population. In some cases, the sample is actually equal to the entire population (e.g., index funds). In other cases, e.g., value or growth mutual funds, the fund is only a sample of the entire population (restricted in this case to a particular class of stocks such as value or growth stocks).

Testing for dangerous chemicals in a reservoir used for drinking water can also make use of the law of large numbers. The goal here is to accurately determine the concentration of a dangerous chemical in the reservoir. While the actual concentration is not known, the law of large numbers tells us that if a large number of samples are taken (with the concentration of the offending chemical being measured for each sample), then the average concentration in the samples will approach the actually expected value of the concentration of the given chemical in the entire reservoir. As with the other two examples, it may be necessary to consider subpopulations. For example, the chemical may concentrate differently on the surface of the reservoir versus the bottom, and so, two sets of samples and associated estimations may be needed.

In general, the law of large numbers can be used to estimate the expected value of some characteristic of a population based on a large number of samples from the population. The actual expected value of the characteristic of interest may be known (such as in the car insurance and stock investing examples above) or it may be unknown (as in the dangerous chemical testing example). Further, one can achieve greater granularity by using subpopulations.

18.3 Central Limit Theorem

The central limit theorem was stated and demonstrated via a Python program in Section 12.9.8. It is mentioned briefly here since the theorem is an example of a universal law of mathematics.

The central limit theorem is used extensively in statistics, e.g., hypothesis testing and confidence intervals. It simplifies problems in statistics by allowing one to work with a distribution that is approximately normal and thereby benefit from the well understood properties of the normal distribution.

18.4 Benford's Law

Benford's law (sometimes referred to as the first-digit law, the significant-digit law, Newcomb–Benford's law or the law of anomalous numbers) is a statistical phenomenon that occurs naturally in many different lists of numerical data. In particular, Benford's Law (first noted in 1881 by the astronomer Simon Newcomb) states that if a number is randomly selected from a list of physical constants or statistical data, the probability that the first digit (i.e., the leading or most significant digit) will be a "1" is about 0.301, rather than $\frac{1}{9} \cong .11$ as might be expected if all digits were equally likely (assuming a multi-digit number does not start with 0).

A list of numbers is said to follow Benford's law if the most significant digit $d \in \{1,2,3, \ldots ,9\}$ occurs with probability

$$P(d) = \log_{10}\left(\frac{d+1}{d}\right)$$

Benford's law applies to a wide variety of data sets, including utility bills, street addresses, stock prices, house prices, population numbers, death rates, lengths of rivers, and physical and mathematical constants. Several well-known infinite sequences of numbers satisfy Benford's Law exactly (as the number of terms approaches infinity), e.g., Fibonacci numbers, factorials, powers of 2 and the powers of almost any other number.

Table 58. Distribution of Leading Digit in Powers of 2 and Fibonacci Number

Digit	Number of Occurrences as Leading Digit in Powers of 2	Number of Occurrences as Leading Digit in Fibonacci Numbers	$P(d) = \log_{10}\left(\frac{d+1}{d}\right)$
1	30	30	30.1%
2	16	18	17.6%
3	13	13	12.5%
4	9	9	9.7%
5	7	9	7.9%
6	7	6	6.7%
7	6	5	5.8%
8	4	6	5.1%
9	5	5	4.6%

The journal article by Collins [88] explains how to apply Microsoft Excel to a data set and determine whether it fits Benford's law. As an example, Table 58 shows the distribution of leading digits for the first 100 powers of 2 and for the first 100 Fibonacci numbers. As one can see, with only 100 data points, the fit to Benford's law (right-hand column of the table) is already fairly good for both sets of numbers.

Benford's law is used to detect possible fraud in fabricated accounting data. This is based on the idea that those who fabricate data (and who do not know about Benford's law) tend to distribute the leading digits (as well as non-leading digits) uniformly, and so an analysis of the first-digit frequency distribution from the questionable data with the expected distribution according to Benford's Law ought to show a suspicious distribution of the leading digit. In the United States, evidence based on Benford's law has been admitted in criminal cases at the federal, state, and local levels. In fact, the United States Internal Revenue Service (IRS) uses Benford's law to detect fabricated data in income tax returns.

In order to apply Benford's law, the following criteria are recommended:

- The data set should evenly span several orders of magnitude, e.g., numbers in the tens, hundreds, thousands and tens of thousands.
- The numbers in the data set should have an equal opportunity to start with any of the digits from 1 to 9. For example, the month of birth (1-12) for famous American politicians is biased since four months start with a 1.

In terms of an explanation for Benford's law, the following statement from the article by A. Berger and T.P. Hill [89] provides a good summary of the current status:

> A broad and often ill-understood phenomenon need not always be reduced to a few theorems. Although many facets of BL *(Benford's Law)* now rest on solid ground, there is currently no unified approach that simultaneously explains its appearance in dynamical systems, number theory, statistics, and real-world data. In that sense, most experts seem to agree with [93] that the ubiquity of BL, especially in real-life data, remains mysterious.

For further details, the reader is referred to the following videos, books and journal articles:

- YouTube video entitled "Number 1 and Benford's Law – Numberphile" [90]
- Benford's Law: Theory and Applications [91]
- An Introduction to Benford's Law [92]
- A Simple Explanation of Benford's Law [93]
- The First Digit Phenomenon [94].

18.5 Power Laws

18.5.1 Overview

A power law is a relationship between two variables, where a change in one variable results in a proportional change in the other variable, independent of the initial size of those quantities, i.e., one variable varies as a power of another.

The simplest case of a power law is a linear relationship between two variables x and y, i.e., a straight line $y = f(x) = ax$ where a is a constant.

The general case is of the form $y = f(x) = \beta x^{-\alpha}$. If x is multiplied by a constant c then y is given by $f(cx) = \beta(cx)^{-\alpha} = c^{-\alpha}(\beta x^{-\alpha}) = c^{-\alpha} f(x)$, i.e., y is changed by a multiplicative factor of $c^{-\alpha}$. Regardless of the value of x, if x is multiplied by c, then y is changed by $c^{-\alpha}$.

For example, if x is the length of a side of a cube and y is the volume, then $y = f(x) = x^3$. So, $\alpha = -3$ and $\beta = 1$. For instance, if the side is increased from 1 to 3 units then the volume is increased by a factor of $3^3 = 27$, and if the side is increased from 2 to 6 units (again, by a factor of 3), then the volume is increased by a factor of 27 (i.e., from 8 to 8*27=216).

Recall the exponential distribution from Section 12.9.4.3. This is an example of a power law where the exponent is negative.

The general form of the power law, i.e., $f(x) = \beta x^{-\alpha}$, can be normalized to define a probability distribution $p(x)$ that is more general than the exponential distribution defined previously. First, we let x_m be the minimum value of x for the distribution. For $x < x_m$, $p(x) = 0$. Next, we determine β to make the area under $p(x)$ equal to 1 (which needs to be the case for a valid PDF).

$$1 = \int_{x_m}^{\infty} p(x)\,dx = \beta \int_{x_m}^{\infty} x^{-\alpha}\,dx = \frac{\beta}{1-\alpha}[0 - x_m^{-\alpha+1}]$$

which implies that

$$\beta = (\alpha - 1)x_m^{\alpha-1}$$

$$p(x) = (\alpha - 1)x_m^{\alpha-1}x^{-\alpha} = \frac{\alpha - 1}{x_m}\left(\frac{x}{x_m}\right)^{-\alpha}$$

The above analysis is under the assumption that $\alpha > 1$; otherwise, the definite integral diverges to negative infinity as $x \to \infty$.

The expected value is given by

$$\mu = E(X) = \int_{x_m}^{\infty} xp(x)\,dx = \beta \int_{x_m}^{\infty} x^{-\alpha+1}\,dx = \frac{\alpha - 1}{\alpha - 2}x_m$$

The computation for the expected value holds true for $\alpha > 2$; otherwise, the definite integral diverges.

The variance is given by $\sigma^2 = \frac{\alpha-1}{\alpha-3}x_m^2$ and is well defined, provided $\alpha > 3$.

The following subsections provide several examples of the power law.

18.5.2 Zipf's Law

Zipf's law is a pattern of distribution in certain data sets, e.g., the words in a language, by which the frequency of occurrence of an item in the data set is inversely proportional to its ranking by frequency. In such a distribution, frequency declines sharply as rank number increases, i.e., a small number of items occur very frequently and a large number of items occur very rarely.

The law was first recognized with respect to linguistics. If one analyzes a large volume of text (e.g., a large book or an entire language), the most common word occurs twice as often as the next most frequent word, three times as often as the next most frequent word after that and so on. The law is named after American linguist George Kingsley Zipf (1902–50) who studied the frequency of different words in the English language.

Amazingly, Zipf's law holds for all languages, even artificially created ones such as Esperanto. Further, Zipf's law holds for many other data sets unrelated to language, such as the population ranks of cities in various countries, corporation sizes and income rankings.

Associated with Zipf's law is the Zipfian distribution, whose PDF is defined as follows:

$$f(x;\alpha,N) = \frac{1/x^{\alpha}}{\sum_{i=1}^{N} 1/i^{\alpha}}$$

where N is the total number of items in the data set, x is the rank and α is a variable used to characterize the distribution. Be warned, however, that many popular explanations of Zipf's law state the relationship between frequency and rank as $f(x) = 1/x$ or $f(x) = c/x$ where c is a constant, neither of which is completely accurate.

There is an excellent YouTube presentation on the topic entitled "The Zipf Mystery" (see https://youtu.be/fCn8zs912OE). Just below the YouTube video is an extensive list of online resources related to Zipf's law.

18.5.3 Pareto Principle

The Pareto principle (often referred to as the 80/20 rule) states that, for many situations, roughly 80% of the effects (good or bad) come from 20% of the causes. The principle is named after economist Vilfredo Pareto, who (back in 1896) noted that approximately 80% of the land in Italy and other countries in his study was owned by 20% of the population. There are many examples of where the law applies, e.g.,

- Software development:
    - 80% of a software product can be written in 20% of the total allocated time.
    - The hardest 20% of the code takes 80% of the time.
    - 80% of software problems are caused by 20% of the bugs.
    - 80% of customers only use 20% of software features.
- Project Management:
    - 80% of the problems come from 20% of the projects.
    - 80% of work is completed by 20% of the associated team.
- Sales and Marketing:
    - 80% of sales come from 20% of the customers.
    - 80% of sales come from 20% of the products.
    - 80% of sales are generated by 20% of the salespeople.
    - 80% of the complaints come from 20% of the customers.

The 80/20 ratio is not exact and a more accurate statement of the principle would be to the effect "the vast majority of effects come from a small minority of the causes."

More formally, there is a probability distribution associated with the Pareto principle (known as, of all things, the Pareto distribution). The PDF for a random variable X following the Pareto distribution is given by

$$f(x) = \begin{cases} \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, & x \geq x_m \\ 0, & x < x_m \end{cases}$$

where x_m is the minimum possible value for X and α is a shape parameter.

Figure 114 shows the PDF for several Pareto random variables, i.e., $x_m = 1$ and $\alpha = 1,2,3$.

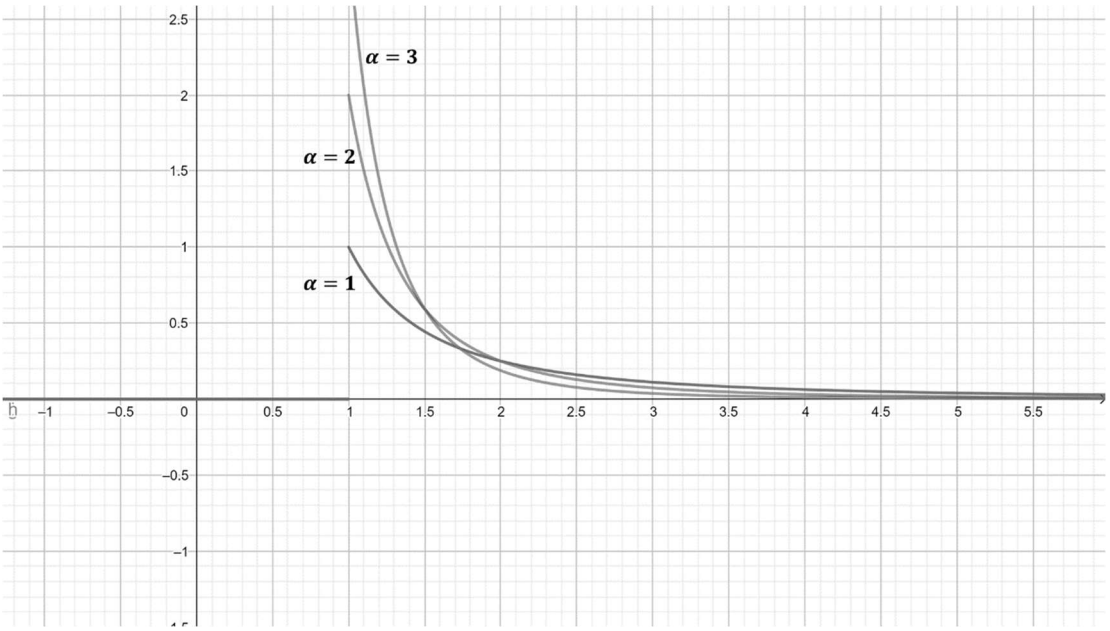


Figure 114. Pareto Probability Distribution Functions

The Pareto distribution coincides with the Pareto principle only when $\alpha \cong 1.16$ (see the Wikipedia article on this topic [95]). So, the Pareto distribution is more general than the Pareto principle.

In general, $\alpha = \frac{\ln(c/100)}{\ln(c/d)}$ implies a $d{:}\,c$ rule. This formula is known as Pareto's Index [96]. For example, to get a 90:10 rule from the Pareto distribution, set $\alpha = \frac{\ln(.10)}{\ln\left(\frac{1}{9}\right)} = 1.048$.

19 Conclusion

The preceding sections have provided a summary of many areas of mathematics with the intent of developing the readers' abilities to think mathematically.

In terms of further learning:

- Wikipedia is a useful source for further exploration and has been referenced extensively in this book.
- The Khan Academy provides free lessons on many topics, including mathematics (see https://www.khanacademy.org/math). In the same vein as the Khan Academy is School Yourself which focuses exclusively on mathematics lessons (see http://schoolyourself.org).
- There is a plethora of YouTube videos on just about any topic in mathematics that can be imagined.
- The Open Culture site provides a list of references to free mathematics textbooks (see http://www.openculture.com/free-math-textbooks).
- MIT Open Courseware offers access to course material and videos related to many MIT courses (including mathematics classes), see https://ocw.mit.edu.

For the readers who have persevered in reading this book, there is the added bonus of now being in a position to better access and understand many interesting topics, including mathematics puzzle books, science magazines articles that make use of mathematics and various popular science books that assume the reader has a background in basic mathematics. Just a small sampling of my favorites:

> Quanta Magazine has many articles that are focused on mathematics (see https://www.quantamagazine.org).
>
> Plus is an online magazine which aims to introduce readers to the beauty and the practical applications of mathematics (see https://plus.maths.org).
>
> Raymond Smullyan has published a wonderful collection of books that teach the reader about formal logic via a collection of puzzles. For a complete list of his books, see https://en.wikipedia.org/wiki/Raymond_Smullyan.
>
> Clifford A. Pickover is a prolific author whose books cross many areas, with a strong focus on mathematics. For a list of his books, see the Wikipedia article: https://en.wikipedia.org/wiki/Clifford_A._Pickover.

There has never been a better time for learning in general with so many online (and in many cases free) textbooks, videos and online courses. Don't let your mathematics learning journey stop with this book!

Acronyms and Symbols

$a \Rightarrow b$ – statement a implies statement b (also written as "if a then b")

$a \Leftrightarrow b$ – statement a implies b, and statement b implies a (also written as "a if and only if b")

$a \vee b$ – the proposition "a or b"

$a \wedge b$ – the proposition "a and b"

$\neg a$ – the proposition "not a"

$\forall$ – for every

$\exists$ – there exists

$\ni$ – such that

$x \in A$ – x is an element of set A

$A \subset B$ – set A is a subset of set B

$A \cup B$ – the union of sets A and B

$A \cap B$ – the intersection of sets A and B

ϕ – the empty set (sometimes written as { })

C_n – cyclic graph of order n

K_n – complete graph of order n

$f'(x)$ – the first derivative of the function $f(x)$

$\int_{x=a}^{x=b} h(x)\,dx$ – definite integral of $h(x)$ from $x = a$ to $x = b$

$\int f(x)\,dx$ – the anti-derivative of the function $f(x)$ with respect to the variable x

$\mathbb{N}$ – the set of natural numbers, i.e., {0, 1, 2, 3, …}

Ω – universal set within a domain of interest

$\mathbb{Q}$ – the set of rational numbers, i.e., both positive and negative fractions and natural numbers

$\mathbb{R}$ – the set of all real numbers

$\mathbb{R}^n$ – the set of points in n-dimensional real space

σ – standard derivation

σ^2 – variance

$\sum_{i=1}^{n} g(i)$ – shorthand for the sum $g(1) + g(2) + \cdots + g(n)$

μ – expected value

$\mathbb{Z}$ – the set of all integers, i.e., all positive and negative whole numbers, and zero

AB – At Bat (baseball stat)

AI – Artificial Intelligence

CDF – Cumulative Density Function

ECP – Equivalence Class Partitioning

GCD – Greatest Common Divisor

GPA – Grade Point Average

IC – Integrated Circuit

iff – if and only if

LCM – Least Common Multiple

ML – Machine Learning

NAND – Not AND

NOR – exclusive OR, i.e., A or B but not both A and B

PDF – Probability Density Function

STEM – Science, Technology, Engineering and Mathematics

wlog – without loss of generality

wrt – with respect to

References

[1] Definition of mathematics from the Encyclopaedia Britannica website, www.britannica.com/science/mathematics, accessed on 7 September 2019.

[2] Euclid's Elements, Wikipedia, en.wikipedia.org/wiki/Euclid's_Elements, accessed on 1 March 2020.

[3] Propositional Logic, article from the Internet Encyclopedia of Philosophy, www.iep.utm.edu/prop-log/, accessed on 13 November 2019.

[4] Sheffer, H. M., 1913, "A set of five independent postulates for Boolean algebras, with application to logical constants", Transactions of the American Mathematical Society, 14: 481–488

[5] Logical connective, Wikipedia, en.wikipedia.org/wiki/Logical_connective, accessed on 15 November 2019.

[6] Levitz, K., Levitz, H., (1979), "Logic and Boolean Algebra", Barron's Educational Series.

[7] NAND Logic, Wikipedia, en.wikipedia.org/wiki/NAND_logic, accessed on 15 November 2019.

[8] Byerly, T.R., 2017, "Introducing Logic and Critical Thinking", Baker Academic.

[9] Gardner, M., 1991, "Chapter 1: The Paradox of the Unexpected Hanging" in the book "The Unexpected Hanging and Other Mathematical Diversions", Chicago University Press, pp. 11-23.

[10] Chow, T. Y., 1998, "The surprise examination or unexpected hanging paradox". The American Mathematical Monthly. 105 (1): 41–51.

[11] Paradox, Wikipedia, en.wikipedia.org/wiki/Paradox, accessed on 20 November 2019.

[12] Quine, W.V., 1976, "The Ways of Paradox, and Other Essays", Harvard University Press; 2nd edition.

[13] Smullyan, Raymond, 1978, "What is the Name of this Book? The Riddle of Dracula and Other Logical Puzzles" (see the section entitled "The Drinking Principle"), Prentice Hall, pp. 209-211.

[14] Universal Set, Wikipedia, en.wikipedia.org/wiki/Universal_set, accessed on 21 November 2019.

[15] Venn, John, 1880, "On the Diagrammatic and Mechanical Representation of Propositions and Reasonings," The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 5. 10 (59): 1–18.

[16] Cardinality, Wikipedia, en.wikipedia.org/wiki/Cardinality, accessed on 30 October 2019.

[17] Halmos, Paul R., 2017, Naïve Set Theory, Dover Publications, Inc. (unabridged republication of the work originally published in 1960 by the D. Van Nostrand Company, Incorporated).

[18] Hrbacek, K., Jech, T., 1999, Introduction to Set Theory, 3rd edition, Marcel Dekker, Inc.

[19] Equivalence partitioning, Wikipedia, en.wikipedia.org/wiki/Equivalence_partitioning, accessed on 14 December 2019.

[20] Smith, Henry J.S., 1874, "On the integration of discontinuous functions," Proceedings of the London Mathematical Society. First series. 6: 140–153.

[21] Cantor, Georg, 1883, "Über unendliche, lineare Punktmannigfaltigkeiten V" [On infinite, linear point-manifolds (sets), Part 5], Mathematische Annalen (in German), 21: 545–591.

[22] Belcastro, Sarah-Marie, Green, Michael, 2001, "The Cantor set contains $\frac{1}{4}$ Really?", The College Mathematics Journal, 32 (1): 55.

[23] Schröder–Bernstein_theorem, Wikipedia, en.wikipedia.org/wiki/Schröder–Bernstein_theorem, accessed 11 November 2019.

[24] Set (mathematics), Wikipedia, en.wikipedia.org/wiki/Set_(mathematics)#cite_note-11, accessed 4 November 2019.

[25] Cantor, Georg, "Beiträge zur Begründung der transfiniten Mengenlehre," Mathematische Annalen, Pages 481 – 512, Berlin, Göttingen, Heidelberg; 1869.

[26] Zermelo–Fraenkel set theory, Wikipedia, en.wikipedia.org/wiki/Zermelo–Fraenkel_set_theory, accessed on 5 November 2019.

[27] Rubin, H. and Rubin, J.E., 1985, "Equivalents of the Axiom of Choice, II", Elsevier Science Publishers B.V.

[28] Suppes, P. 2012, "Axiomatic Set Theory", Dover Publications.

[29] Supertasks, Wikipedia, en.wikipedia.org/wiki/Supertask, accessed 12 November 2019.

[30] Solomon, A., 1990, "The Essentials of Boolean Algebra", Research and Education Association.

[31] Whitesitt, J.E., 2010, "Boolean Algebra and Its Applications", Dover Publications.

[32] Polynomial, Wikipedia, en.wikipedia.org/wiki/Polynomial, accessed 30 November 2019.

[33] Geometrical properties of polynomial roots, Wikipedia, en.wikipedia.org/wiki/Geometrical_properties_of_polynomial_roots, accessed on 30 November 2019.

[34] Well-ordering theorem, Wikipedia, en.wikipedia.org/wiki/Well-ordering_theorem, accessed on 10 December 2019.

[35] Graham, R., Knuth, D., Patashnik, O., 1994, "Concrete Mathematics: A Foundation for Computer Science", 2nd edition, Addison-Wesley.

[36] Divisibility, Wikipedia, en.wikipedia.org/wiki/Divisibility_rule, accessed on 17 December 2019.

[37] Euclid's Elements, Book IX, Proposition 20, translation into English by David E. Joyce (1996), mathcs.clarku.edu/~djoyce/java/elements/bookIX/propIX20.html, accessed 9 July 2020.

[38] Tanton, J., (2010), "THINKING MATHEMATICS 2: Advanced Counting and Advanced Algebra Systems", self-published by the author (available at www.jamestanton.com).

[39] List of limits, Wikipedia, en.wikipedia.org/wiki/List_of_limits, accessed on 1 January 2020.

[40] e (mathematical constant), Wikipedia, en.wikipedia.org/wiki/E_(mathematical_constant), accessed on 15 March 2020.

[41] Differentiation rules, Wikipedia, en.wikipedia.org/wiki/Differentiation_rules, accessed on 31 December 2019.

[42] Poker probabilities, Wikipedia, en.wikipedia.org/wiki/Poker_probability, accessed 5 December 2019.

[43] Odds, Wikipedia, en.wikipedia.org/wiki/Odds#Fractional_odds, accessed on 21 December 2019.

[44] Inclusion-exclusion Principle, Wikipedia, en.wikipedia.org/wiki/Inclusion–exclusion_principle, accessed 22 December 2019.

[45] Measure (mathematics), Wikipedia, en.wikipedia.org/wiki/Measure_(mathematics), accessed on 22 December 2019.

[46] Hogg, R.V., Craig, A.T., 1978, Introduction to Mathematical Statistics, Macmillan Publishing Co., Inc.

[47] Arias, E., Xu, J., “United States Life Tables, 2017”, National Vital Statistics Reports, Volume 68, Number 7.

[48] Bernardo, J.M., Smith, A.F.M., 2000, Bayesian Theory, John Wiley and Sons, Ltd.

[49] Davidson-Pilon, C., 2016, Bayesian Methods for Hackers, Addison-Wesley.

[50] Donovan, T.M., Mickey, R.M., 2019, Bayesian Statistics for Beginners, Oxford University Press.

[51] National Weather Service: Wind Chill Chart, www.weather.gov/safety/cold-wind-chill-chart, accessed on 29 December 2019.

[52] List of probability distributions, Wikipedia, en.wikipedia.org/wiki/List_of_probability_distributions, accessed on 2 January 2020.

[53] Linearity of Expectation Function, Proof Wiki, proofwiki.org/wiki/Linearity_of_Expectation_Function, accessed on 4 January 2020.

[54] Variance of Linear Combination of Random Variables, Proof Wiki, proofwiki.org/wiki/Variance_of_Linear_Combination_of_Random_Variables, accessed 17 March 2020.

[55] Variance of Binomial Distribution, Proof Wiki, proofwiki.org/wiki/Variance_of_Binomial_Distribution, accessed on 4 January 2020.

[56] Standard score, Wikipedia, en.wikipedia.org/wiki/Standard_score, 17 March 2020.

[57] H. Pishro-Nik, 2014, "Introduction to probability, statistics, and random processes", available at https://www.probabilitycourse.com, Kappa Research LLC.

[58] Hoel, P.G., Port, S.C., Stone, C.J., 1971, Introduction to Probability Theory, Houghton Mifflin Company.

[59] M/M/1 queue, Wikipedia, en.wikipedia.org/wiki/M/M/1_queue, accessed on 5 January 2020.

[60] Marcus, A.H., 1965, “On a Test of Randomness of Lunar Craters”, Astronomical Journal, Vol. 70, p. 325. Abstract available at https://ui.adsabs.harvard.edu/abs/1965AJ.....70..325M/abstract.

[61] Proof that the Binomial Distribution tends to the Poisson Distribution, YouTube, youtu.be/ceOwlHnVCqo, accessed 5 July 2020.

[62] Bessel's correction, Wikipedia, en.wikipedia.org/wiki/Bessel's_correction, accessed 13 January 2020.

[63] Confidence interval, Wikipedia, en.wikipedia.org/wiki/Confidence_interval, accessed on 3 August 2020.

[64] Stephanie Glen, "T-Score vs. Z-Score: What’s the Difference?" from StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/t-score-vs-z-score/, accessed on 15 June 2020.

[65] Student’s t-distribution, Wikipedia, en.wikipedia.org/wiki/Student's_t-distribution#Table_of_selected_values, accessed on 22 January 2020.

[66] Test statistic, Wikipedia, en.wikipedia.org/wiki/Test_statistic, accessed 16 January 2020.

[67] Student’s t-test, Wikipedia, en.wikipedia.org/wiki/Student%27s_t-test, accessed 18 January 2020.

[68] Hypothesis Test vs. Confidence Interval | Statistics Tutorial #15 | MarinStatsLectures, YouTube, https://youtu.be/J-yMiTaai4c, accessed 20 July 2020.

[69] Simple linear regression, Wikipedia, en.wikipedia.org/wiki/Simple_linear_regression, accessed 23 January 2018.

[70] Berkson's paradox, Wikipedia, en.wikipedia.org/wiki/Berkson%27s_paradox, accessed 30 January 2020.

[71] Planar graph, Wikipedia, en.wikipedia.org/wiki/Planar_graph, accessed on 20 March 2020.

[72] Polya, G., Conway, J.H., 2014, “How to Solve It”, Princeton University Press.

[73] Wohlgemuth, A., 2011, “Introduction to Proof in Abstract Mathematics”, Dover Publications, Inc.

[74] Sundstrom, T., 2019, “Mathematical Reasoning: Writing and Proof”, Version 2.1, self-published by author, available at www.tedsundstrom.com/mathematical-reasoning-writing-and-proof.

[75] Reductio ad absurdum, Wikipedia, en.wikipedia.org/wiki/Reductio_ad_absurdum, accessed on 9 March 2020.

[76] Wiles, A., 1995, “Modular elliptic curves and Fermat’s Last Theorem”, Annals of Mathematics, 142, p.443-551.

[77] Wiles's proof of Fermat's Last Theorem, Wikipedia, en.wikipedia.org/wiki/Wiles's_proof_of_Fermat's_Last_Theorem, accessed on 4 March 2020.

[78] List of lemmas, Wikipedia, en.wikipedia.org/wiki/List_of_lemmas, accessed on 17 June 2020.

[79] Definition of algorithm from “The Definitive Glossary of Higher Mathematical Jargon”, Math Vault, mathvault.ca/math-glossary/#algo, accessed 22 February 2020.

[80] Tower of Hanoi, Wikipedia, en.wikipedia.org/wiki/Tower_of_Hanoi, accessed 12 February 2020.

[81] Parallel algorithms for minimum spanning trees, Wikipedia, en.wikipedia.org/wiki/Parallel_algorithms_for_minimum_spanning_trees, accessed 12 February 2020.

[82] Grid computing, Wikipedia, en.wikipedia.org/wiki/Grid_computing, accessed 12 February 2020.

[83] Akl, S.G., (1985), “Parallel Sorting Algorithms”, Academic Press, Inc.

[84] General number field sieve, Wikipedia, en.wikipedia.org/wiki/General_number_field_sieve#Implementations, accessed 13 February 2020.

[85] ISO 5807:1985, “Information processing — Documentation symbols and conventions for data, program and system flowcharts, program network charts and system resources charts”, International Organization for Standardization.

[86] Nassi–Shneiderman diagram, Wikipedia, en.wikipedia.org/wiki/Nassi–Shneiderman_diagram, accessed on 13 February 2020.

[87] DRAKON, Wikipedia, en.wikipedia.org/wiki/DRAKON, accessed on 13 February 2020.

[88] Collins, C.J., April 2017, “Using Excel and Benford’s Law to detect fraud”, Journal of Accountacy, www.journalofaccountancy.com/issues/2017/apr/excel-and-benfords-law-to-detect-fraud.html.

[89] Berger, A., Hill, T.P., Benford’s Law Strikes Back: No Simple Explanation in Sight for Mathematical Gem, Math Intelligencer 33, 85–91 (2011).

[90] Number 1 and Benford's Law – Numberphile, YouTube, youtu.be/XXjlR2OK1kM, accessed on 27 July 2020.

[91] Miller, S.J. (editor), 2015, “Benford’s Law: Theory and Applications”, Princeton University Press.

[92] Berger, A., Hill, T.P., 2015, “An Introduction to Benford’s Law”, Princeton University Press.

[93] Fewster, R.M., “A Simple Explanation of Benford’s Law”, The American Statistician, February 2009, Vol. 63, No. 1.

[94] Hill, T.P., “The First Digit Phenomenon”, American Scientist, July-August 1998, Vol. 86, Issue 4, p. 358.

[95] Pareto distribution, Wikipedia, en.wikipedia.org/wiki/Pareto_distribution, accessed on 22 February 2020.

[96] Pareto’s index, Wikipedia, en.wikipedia.org/wiki/Pareto_index, accessed on 18 June 2020.

Index of Terms

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

Z