MATHEMATICS IN CYBERSECURITY

ALFRED BASTA · STEPHAN DELONG
STAVROS BASTA



WILEY

Mathematics in Cybersecurity

Mathematics in Cybersecurity

Alfred Basta
Stephan DeLong
Stavros Basta

WILEY

Copyright © 2026 by John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

The manufacturer's authorized representative according to the EU General Product Safety Regulation is Wiley-VCH GmbH, Boschstr. 12, 69469 Weinheim, Germany, e-mail: Product_Safety@wiley.com.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and the authors have used their best efforts in preparing this work, including a review of the content of the work, neither the publisher nor the authors make any representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data Applied for:

Hardback ISBN: 9781394304769

Cover Design: Wiley
Cover Image: © Xuanyu Han/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

Alfred Basta
In loving memory of my devoted wife and coauthor, Nadine, whose unwavering faith and boundless encouragement formed the bedrock of my journey into the realm of cybersecurity. Your steadfast belief in my abilities ignited and sustained my passion for this field, and your presence beside me made every challenge conquerable. Though you are no longer here to witness this milestone, your spirit lives on in every page.

To my beloved daughter, Rebecca, you are the radiant light that illuminates my path forward. Your insatiable curiosity and limitless imagination serve as daily reminders of why we must continue to push beyond conventional boundaries and venture into uncharted territories. May this work stand as a testament to your infinite potential, and may you forever possess the courage to chase your most ambitious dreams.

To my cherished son, Stavros, your infectious enthusiasm and remarkable tenacity have been the driving forces behind my every pursuit. Through you, I have learned the profound meaning of perseverance and discovered the transformative power of embracing challenges with open arms. As you continue to grow and evolve, let this book remind you that with unwavering determination and resilience, no goal is beyond your reach.

In eternal remembrance of my dear parents, who planted within me the seeds of diligence, determination, and unwavering perseverance. They were the guiding stars of my youth and the inspiration that set me upon this remarkable journey. Though their physical presence has passed beyond our reach, their love, wisdom, and enduring support continue to illuminate my way forward, today and always.

Stephan DeLong
I dedicate this book to the memory of my mother, Bonita J. DeLong, to the kindness, skill, and friendship of Dr. Alfred Basta, to my wife Kathryn Farber DeLong, and to my "bestie," a tripod cancer-surviving cat, Jojo DeLong.

Stavros Basta
To my father, Dr. Alfred Basta, whose unwavering support made this book possible and who has been my beacon throughout my academic journey—your wisdom and encouragement have shaped not only my career but also my character.

To the cherished memory of my mother, who continues to be my guiding light even in her absence—your love and values illuminate every path I choose.

And to Rebecca Basta, whose tireless help, endless patience, and steadfast support have been invaluable throughout this endeavor—this work would not exist without your dedication. This book stands as a testament to the power of family, love, and the profound impact of those who believe in us.

Contents

About the Authors *xix*
Acknowledgments *xxi*
About the Companion Website *xxiii*

1 Introduction to Sets *1*
1.1 Introduction *1*
1.1.1 The Definition of a Set *2*
1.1.2 The Elements of a Set *2*
1.1.3 Cardinality of a Set *3*
1.1.4 The Notation of Sets *3*
1.2 Types of Sets *5*
1.2.1 The Empty Set (The Null Set) *5*
1.2.2 Singleton Sets *6*
1.2.3 Finite and Infinite Sets *6*
1.2.4 Special Mathematical Sets *7*
1.2.5 Subset and Superset *7*
1.2.6 Power Set *8*
1.3 Operations on Sets *8*
1.3.1 Union of Sets *8*
1.3.2 Intersection of Sets *9*
1.3.3 Cardinality of Unions and Intersections *10*
1.3.4 Difference of Sets *10*
1.4 Set Relations *11*
1.4.1 Equivalence of Sets *12*
1.4.2 Equivalence and Non-equivalence of Infinite Sets *12*
1.4.3 Disjoint Sets *13*
1.4.4 Universal Sets *13*
1.4.5 The Complement of a Set *14*
1.4.6 Venn Diagrams *14*

2 Logic *17*
2.1 Introduction to Logic *17*
2.1.1 Definition of Logic *17*
2.1.2 Importance of Logic in Discrete Mathematics *18*
2.1.3 Propositions and Logical Connectives *18*

2.2 Propositional Logic *19*
2.2.1 Propositions and Compound Propositions *19*
2.2.2 Logical Connectives and Truth Values *20*
2.2.2.1 Negation (NOT) *20*
2.2.2.2 Conjunction (AND) *20*
2.2.2.3 Disjunction (OR) *21*
2.2.2.4 Implication (IF–THEN) *21*
2.2.2.5 Biconditional (IF AND ONLY IF) *21*
2.2.3 Truth Tables *22*
2.2.3.1 Construction of Truth Tables *22*
2.2.3.2 Categorization of Truth Table Results *23*
2.2.3.3 Tautology, Contradiction, and Contingency *24*
2.3 Logical Equivalence *26*
2.3.1 Definition of Logical Equivalence *26*
2.3.2 Laws of Logical Equivalence *27*
2.3.2.1 Commutative Laws of Conjunction and Disjunction *27*
2.3.2.2 Associative Laws *27*
2.3.2.3 Distributive Laws *28*
2.3.2.4 De Morgan's Laws *29*
2.3.2.5 Absorption Laws *29*
2.3.2.6 Idempotent Laws *29*
2.3.2.7 Double Negation Law *30*
2.3.2.8 Simplification Law *30*
2.4 Logical Inference and Proofs *30*
2.4.1 Validity and Soundness *30*
2.4.2 Modus Ponens and Modus Tollens *31*
2.4.3 Hypothetical Syllogism *32*
2.4.4 Disjunctive Syllogism *32*
2.4.5 Constructing Proofs Using Inference Rules *32*
2.5 Logic Gates *33*
2.5.1 Introduction to Logic Gates *33*
2.5.2 Types of Logic Gates *34*
2.5.2.1 NAND Gate *34*
2.5.2.2 NOR Gate *35*
2.5.2.3 XOR Gate *35*
2.5.2.4 XNOR Gate *36*
2.5.3 Combinational Logic Circuits *36*
2.5.4 Sequential Logic Circuits *38*
2.6 Predicate Logic *40*
2.6.1 Predicates and Quantifiers *40*
2.6.2 Universal Quantifier *40*
2.6.3 Existential Quantifier *41*
2.6.4 Negation of Quantified Statements *41*
2.6.5 Translating English Sentences into Predicate Logic *42*
2.7 Applications of Logic *43*
2.7.1 Digital Systems and Boolean Algebra *43*
2.7.2 Computer Programming and Logical Operations *45*

2.7.3 Artificial Intelligence and Logical Reasoning *46*
2.7.4 Philosophical Arguments and Logical Fallacies *47*
2.8 Summary *49*

3 Introduction to Number Systems *51*
3.1 What Is a Number System? *51*
3.1.1 Importance of Number Systems in Discrete Math and Computer Science *52*
3.1.2 Place Value and Positional Notation in a Number System *52*
3.2 Binary Number System *52*
3.2.1 Introduction to Binary Numbers *52*
3.2.2 Binary Place Value and Positional Notation *53*
3.2.3 Converting Decimal Numbers to Binary *54*
3.2.3.1 Division Method *54*
3.2.3.2 Subtraction Method *55*
3.2.4 Converting Binary Numbers to Decimal *56*
3.2.5 Binary Arithmetic *56*
3.2.5.1 Binary Addition *56*
3.2.5.2 Binary Subtraction *57*
3.2.5.3 Binary Multiplication *59*
3.2.5.4 Binary Division *59*
3.2.6 Signed Binary Numbers *61*
3.2.6.1 Sign-Magnitude Representation *62*
3.2.6.2 Two's Complement Representation *63*
3.3 Octal (Base-8) Number System *64*
3.3.1 Introduction to Octal Numbers *64*
3.3.2 Converting Decimal to Octal *65*
3.3.3 Converting Octal to Decimal *65*
3.3.4 Converting Binary to Octal *66*
3.3.5 Converting Octal to Binary *67*
3.3.6 Addition and Subtraction in Octal *68*
3.3.7 Multiplication and Division in Octal *70*
3.4 Hexadecimal (Hex) Number System *75*
3.4.1 Introduction to Hexadecimal Numbers *76*
3.4.2 Converting Decimal to Hexadecimal *76*
3.4.3 Converting Hexadecimal to Decimal *77*
3.4.4 Converting Binary to Hexadecimal *78*
3.4.5 Converting Hexadecimal to Binary *79*
3.5 Ternary Number System *81*
3.5.1 Balanced Ternary Number System *82*
3.6 Generalization: Base-n Number Systems *84*
3.7 Applications of Number Systems *85*
3.7.1 Number Systems in Computer Architecture *85*
3.7.1.1 Binary Number System *85*
3.7.1.2 Hexadecimal Number System *85*
3.7.1.3 Octal Number System *86*
3.7.2 Number Systems in Digital Logic Design *86*
3.7.2.1 Binary Number System *86*

3.7.2.2 Hexadecimal Number System *87*
3.7.3 Number Systems in Cryptography *88*
3.7.3.1 Binary Number System *88*
3.7.3.2 Modular Arithmetic and Prime Numbers *88*
3.8 Arithmetic in Other Number Systems *89*
3.8.1 Addition and Subtraction in Hexadecimal *89*
3.8.1.1 Addition in Hexadecimal *89*
3.8.1.2 Subtraction in Hexadecimal *90*
3.8.2 Multiplication and Division in Hexadecimal *93*
3.8.2.1 Multiplication in Hexadecimal *93*
3.8.2.2 Division in Hexadecimal *94*
3.9 Floating-Point Representation *96*
3.9.1 Fixed-Point vs. Floating-Point Numbers *96*
3.9.1.1 Fixed-Point Numbers *96*
3.9.1.2 Floating-Point Numbers *96*
3.9.2 IEEE 754 Floating-Point Standard *98*
3.9.2.1 IEEE 754 Floating-Point Arithmetic *99*
3.9.2.2 Rounding Modes *100*
3.9.2.3 Single-Precision and Double-Precision Formats *100*

4 Straight-Line Equations and Graphs *103*
4.1 Introduction to Straight Lines *103*
4.2 Definition of a Straight Line *103*
4.3 The Intercepts of a Line *105*
4.4 Slope of a Line *106*
4.4.1 The Rise over Run Interpretation of the Slope of a Line *108*
4.4.2 Slope as a Rate of Change *108*
4.5 The Equation of a Line *109*
4.6 Standard Form of a Straight-Line Equation *112*
4.7 The Primary Uses of the Three Forms of the Equation of a Line *113*
4.8 Graphing Straight Lines *113*
4.9 Graphing Using the Standard Form *114*
4.10 Parallel and Perpendicular Lines *118*
4.11 Applications of Straight-Line Equations: Modeling Linear Relationships *119*
4.11.1 Application: Direct Variation (Also Called Direct Proportionality) and Straight-Line Equations *120*

5 Introduction to Systems of Linear Equations *123*
5.1 Definition and Notation of Systems of Linear Equations *123*
5.2 Graphical Representation of a System of Linear Equations *123*
5.3 The Solution to a System of Two Linear Equations in Two Variables *126*
5.3.1 Checking the Solution to a System of Two Linear Equations in Two Variables *127*
5.3.2 Statement of the Solution to Linear Systems of Equations *128*
5.3.2.1 Consistent System *128*
5.3.2.2 Inconsistent System *128*
5.3.2.3 Dependent System *128*
5.3.3 Solving Systems of Linear Equations by the Substitution Method *128*
5.3.4 Substitution, the Inconsistent Case, and the Dependent Case *133*

5.3.5 Solving Systems of Equations Using the Method of Elimination *134*
5.4 Applications That Lead to Systems of Linear Equations *139*
5.4.1 Systems Involving Three Variables *140*
5.5 Introduction to Matrices *148*
5.5.1 The Fundamentals of Matrices *148*
5.5.2 Definition and Terminology *148*
5.5.3 Matrix Operations *149*
5.5.3.1 Matrix Addition and Subtraction *149*
5.5.3.2 Multiplication of a Matrix by a Scalar *153*
5.5.3.3 Matrix Multiplication *154*
5.5.3.4 Formal Definition of Matrix Multiplication *155*
5.5.3.5 The Multiplicative Inverse of a Square Matrix *158*
5.5.4 Using Matrices to Solve Systems of Linear Equations *165*
5.5.4.1 Representing a System of Linear Equations as a Matrix Equation *165*
5.5.4.2 Augmented Matrix and Gaussian Elimination *167*

6 Sequences and Series *177*
6.1 Sequences *177*
6.1.1 Sequences Given by Explicit Formulas Using the Index as the Independent Variable *178*
6.1.2 Sequences Given by a Recursive Formula *178*
6.1.3 Comparison of Generating Formulas and Recursive Formulas *179*
6.1.4 Sequences Given by Tables *180*
6.1.5 Sequences Given by Graphs of Discrete Points in the Plane *180*
6.2 Arithmetic Sequences *181*
6.2.1 Summation Notation *184*
6.2.1.1 Finding the Sum of Terms of a Finite Series *184*
6.2.1.2 Sum of the First *n* Terms of an Arithmetic Sequence *185*
6.2.1.3 Finding the Sum of the First *n* Terms of an Arithmetic Sequence with Varying Common Differences *187*
6.3 Geometric Sequences *189*
6.3.1 Infinite Geometric Sequences *189*
6.3.2 Convergence of Infinite Geometric Sequences *191*
6.4 Fibonacci Sequence *192*
6.4.1 Properties of the Fibonacci Sequence *192*
6.4.2 Applications of the Fibonacci Sequence *195*
6.4.3 Applications of Fibonacci Sequence in Cybersecurity *196*
6.4.3.1 Cryptography and Encryption *196*
6.4.3.2 Password Generation and Management *196*
6.4.3.3 Network Security and Traffic Analysis *196*
6.5 The Principles of Mathematical Induction *196*
6.6 Factorial Notation and Binomial Coefficients (Combinations) *199*
6.6.1 The Binomial Theorem *200*
6.6.2 Graphical Interpretation and Pascal's Triangle *203*
6.6.3 Applications of the Binomial Theorem *203*
6.6.4 Applications of the Binomial Theorem in Cryptography *203*
6.6.4.1 Modular Arithmetic and the Binomial Theorem *203*
6.6.4.2 The Binomial Theorem and Pseudo-random Number Generation *204*

7 Measuring an Angle *205*
7.1 Measuring an Angle *205*
7.1.1 Degrees *207*
7.1.2 Radians *208*
7.1.3 Conversion Between Degrees and Radians *209*
7.1.4 Measuring Angles *211*
7.2 Trigonometric Functions *214*
7.2.1 Right Triangles *214*
7.2.2 Primary Trigonometric Functions *215*
7.2.3 Secondary Trigonometric Functions *216*
7.2.4 Simplifications That Are Not Exact *216*
7.2.5 Recognizing When a Square Root Can Be Simplified and Performing the Simplification *217*
7.2.6 Fundamental Right Triangles to Be Memorized *217*
7.2.6.1 The 1–2–$\sqrt{3}$ Triangle *217*
7.2.6.2 The 1–1–$\sqrt{2}$ Triangle *218*
7.2.7 Trigonometric Function Values of Any Real Number and Graphing a Trigonometric Function *220*
7.2.8 The Inverse Trigonometric Functions *224*
7.3 Introduction to Trigonometric Identities *226*
7.3.1 Basic Trigonometric Identities *226*
7.3.2 Using the Fundamental Identities in Computations *228*
7.4 Trigonometric Functions of the Sum of Two Angles *230*
7.4.1 Sum and Difference Identities *230*
7.5 Trigonometric Functions of the Difference of Two Angles *232*
7.5.1 Difference Identities *232*
7.6 Trigonometric Functions of Double-Angles *234*
7.6.1 Double-Angle Identities *234*
7.7 Trigonometric Functions of Half-Angles *236*
7.7.1 Half-Angle Identities *236*

8 Introduction to Complex Numbers *241*
8.1 Introduction to Complex Numbers *241*
8.1.1 The Problem to Be Solved *242*
8.1.2 What Are Complex Numbers? *242*
8.1.3 The Complex Plane *242*
8.2 Addition and Subtraction of Complex Numbers *243*
8.2.1 Addition of Complex Numbers *243*
8.2.2 Subtraction of Complex Numbers *244*
8.2.3 Geometric Interpretation of Complex Numbers as Points or Vectors *245*
8.2.4 Example with Geometric Interpretation *247*
8.2.5 The Powers of i *248*
8.3 Multiplication and Division of Complex Numbers *248*
8.3.1 Division of Complex Numbers *249*
8.3.2 Important Algebraic Properties *251*
8.4 Graphical Representation of a Complex Number *251*
8.4.1 Magnitude and Argument *251*

8.4.2 Polar Form of a Complex Number *252*
8.5 Exponential Form *254*
8.5.1 Applications of Complex Numbers *256*
8.6 Applications of Complex Numbers in Cybersecurity *257*

9 Exponentials and Algorithms *259*
9.1 Exponential Functions *259*
9.1.1 Properties of Exponential Functions When Stated in General Form *259*
9.1.2 The Effect of the Value of the Base *260*
9.1.3 The Common Base of Exponentials *261*
9.1.4 The Natural Base of Exponentials *263*
9.1.5 Applications of Exponential Functions *264*
9.2 Logarithms *271*
9.3 The Properties of Exponential and Logarithmic Function Graphs *281*
9.4 Solving Exponential and Logarithmic Equations *284*
9.4.1 Solving Exponential Equations *284*
9.4.2 Solving Logarithmic Equations *286*
9.5 Applications of Exponential and Logarithmic Functions in Cybersecurity *289*
9.5.1 Cryptographic Algorithms *289*
9.5.2 RSA Cryptography *289*
9.5.3 Network Security and Traffic Analysis *289*
9.5.4 Network Traffic Analysis *289*
9.5.5 Vulnerability Assessment and Risk Analysis *290*
9.5.6 Intrusion Detection and Anomaly Identification *291*

10 What Do We Mean by Statistics? *293*
10.1 Sampling *293*
10.1.1 Convenience Sampling *294*
10.1.2 Systematic Sampling *294*
10.1.3 Simple Random Sampling *295*
10.1.4 Cluster Sampling *295*
10.1.5 Stratified Sampling *296*
10.2 Statistical Graphs *297*
10.2.1 Bar Graph *297*
10.2.2 Pie Chart *299*
10.2.3 Line Graph *300*
10.2.4 Scatter Plot *302*
10.2.5 Histogram *303*
10.3 The Measures of Central Tendency *304*
10.3.1 Mean *304*
10.3.2 Median *304*
10.3.3 Mode *304*
10.4 The Measures of Dispersion *307*
10.4.1 Range *307*
10.4.2 Variance *308*
10.4.3 Standard Deviation *309*
10.5 The Normal Distribution *311*

10.5.1 Standard Normal Distribution *312*
10.6 The Binomial Distribution *323*
10.7 Linear Correlation and Regression *329*
10.7.1 Correlation Coefficient *329*
10.7.2 Least Squares Regression Line *331*

11 Basic Concepts of Probability *341*
11.1 Experiments, Outcomes, Sample Spaces, and Events *341*
11.1.1 Theoretical Probability *344*
11.1.2 Empirical Probability *346*
11.2 Axioms of Probability *348*
11.3 Basic Rules of Probability *349*
11.3.1 The Addition Rule for Events That Are NOT Mutually Exclusive *349*
11.3.2 The Multiplication Rule for Independent Events *349*
11.3.3 Conditional Probability *350*
11.3.4 The Multiplication Rule for Dependent Events *351*
11.4 Counting Techniques and Probability *352*
11.4.1 Permutations and Combinations *352*
11.4.2 Permutations *353*
11.4.3 Combinations *354*
11.5 Random Variables *355*
11.6 Probability Mass Function *355*
11.6.1 Graphical Representation of the PMF *356*
11.7 Distributions of Discrete Data *356*
11.7.1 Bernoulli Distribution *357*
11.7.2 Binomial Distribution *358*
11.7.3 Geometric Distribution *358*
11.7.4 Poisson Distribution *360*
11.7.5 Hypergeometric Distribution *362*
11.7.6 Expected Value and Variance of Discrete Random Variables *364*
11.7.7 Variance *366*
11.8 Continuous Probability Distributions *367*
11.8.1 Probability Density Functions *367*
11.8.2 Distributions of Continuous Random Variables *368*
11.8.3 Uniform Distribution *368*
11.8.4 Normal Distribution *369*
11.8.5 Exponential Distribution *370*
11.8.6 Expected Value and Variance of Continuous Random Variables *372*
11.8.6.1 Expected Value of Continuous Random Variables *372*
11.8.6.2 Variance of Continuous Random Variables *373*
11.9 The Central Limit Theorem *374*
11.9.1 Understanding the Central Limit Theorem *374*
11.9.2 Visualizing the Central Limit Theorem *375*

12 Graph Theory *377*
12.1 What Is Graph Theory? *377*
12.1.1 Definition of a Graph *377*

12.1.2 Brief History of Graph Theory *379*
12.1.2.1 Leonhard Euler and the Königsberg Bridge Problem (1736) *379*
12.1.2.2 Gustave Kirchhoff and the Study of Trees (1845) *379*
12.1.2.3 Arthur Cayley and the Enumeration of Trees (1857) *379*
12.1.2.4 William Hamilton and the Icosian Game (1857) *379*
12.1.2.5 Dénes Kőnig and the First Graph Theory Book (1936) *379*
12.1.2.6 Claude Shannon and the Application of Graph Theory to Communication Networks (1948) *380*
12.1.2.7 Paul Erdős and the Rise of Modern Graph Theory (1950s–1990s) *380*
12.1.3 Applications of Graph Theory *380*
12.1.3.1 Computer Science *380*
12.1.3.2 Social Networks *380*
12.1.3.3 Biology *381*
12.1.3.4 Chemistry *381*
12.1.3.5 Other Fields *381*
12.2 Exploration of Basic Concepts *381*
12.2.1 Vertices (Nodes) *381*
12.2.2 Edges *383*
12.2.2.1 Directed Edges *383*
12.2.2.2 Undirected Edges *383*
12.2.3 Loops and Multiple Edges *385*
12.2.3.1 Loops *385*
12.2.3.2 Adjacent Vertices *385*
12.2.3.3 Adjacency Matrix *387*
12.2.3.4 Incident Edges *390*
12.2.3.5 Incidence Matrix *391*
12.2.3.6 Multiple Edges *391*
12.2.3.7 Handling Loops and Multiple Edges *393*
12.2.3.8 Density of a Graph *394*
12.2.3.9 Diameter of a Graph *394*
12.3 Types of Graphs *395*
12.3.1 Undirected Graphs *395*
12.3.2 Directed Graphs (Digraphs) *396*
12.3.3 Weighted Graphs *397*
12.3.4 Simple Graphs *398*
12.3.5 Mixed Graphs *399*
12.3.6 Multigraphs *400*
12.3.7 Complete Graphs *402*
12.3.8 Bipartite Graphs *405*
12.3.9 Regular Graphs *407*
12.4 Adjacency Matrix *408*
12.4.1 Adjacency List *409*
12.4.2 Incidence Matrix *411*
12.4.3 Edge List *411*
12.5 Graph Traversals *413*
12.5.1 Depth-First Search (DFS) *413*
12.5.1.1 Algorithm *413*

12.5.1.2 Applications *415*
12.5.2 Breadth-First Search (BFS) *415*
12.5.2.1 Applications *417*
12.6 Connectivity *418*
12.6.1 Connected Graphs *418*
12.6.2 Strongly Connected Graphs *420*
12.6.3 Weakly Connected Graphs *421*
12.6.4 Components *422*
12.7 Special Graphs *423*
12.7.1 Trees *423*
12.7.1.1 Definition *423*
12.7.1.2 Properties *423*
12.7.2 Planar Graphs *424*
12.7.2.1 Definition *424*
12.7.2.2 Euler's Formula *425*
12.7.3 Eulerian Graphs *427*
12.7.3.1 Definition *428*
12.7.3.2 Necessary and Sufficient Conditions *429*
12.7.3.3 Eulerian and Semi-Eulerian Trails *430*
12.7.4 Hamiltonian Graphs *430*
12.7.4.1 Necessary Conditions *431*
12.7.4.2 Sufficient Conditions *432*
12.8 Graph Algorithms *432*
12.8.1 Shortest Path Algorithms *432*
12.8.1.1 Dijkstra's Algorithm *433*
12.8.1.2 Floyd–Warshall Algorithm *434*
12.8.2 Minimum Spanning Tree Algorithms *436*
12.8.2.1 Kruskal's Algorithm *436*
12.8.2.2 Prim's Algorithm *438*
12.8.3 Network Flow Algorithms *441*
12.8.3.1 Ford–Fulkerson Algorithm *441*
12.8.3.2 Max-Flow Min-Cut Theorem *443*

13 What Is Cryptography? *445*
13.1 What Is Cryptography? *445*
13.1.1 Definition and Purpose *445*
13.1.2 Brief History of Cryptography *446*
13.2 Basic Cryptographic Concepts *446*
13.2.1 Plaintext, Ciphertext, and Keys *447*
13.2.2 Encryption, Decryption, and Hacking *447*
13.2.3 Symmetric and Asymmetric Cryptography *447*
13.3 Classical Ciphers *448*
13.3.1 Substitution Ciphers *448*
13.3.1.1 Caesar Cipher *449*
13.3.1.2 Monoalphabetic Substitution *451*
13.3.1.3 Polyalphabetic Substitution *453*
13.3.2 Transposition Ciphers *453*

13.3.2.1 Rail Fence Cipher *453*
13.3.2.2 Columnar Transposition *455*
13.4 Modern Cryptography *459*
13.4.1 Symmetric Key Cryptography *460*
13.4.1.1 Block Ciphers (e.g., AES) *460*
13.4.1.2 Stream Ciphers (e.g., RC4) *460*
13.4.2 Public Key Cryptography *462*
13.4.2.1 RSA Cryptosystem *462*
13.4.2.2 Diffie-Hellman Key Exchange *464*
13.4.3 Hash Functions and Message Authentication Codes *466*
13.4.3.1 SHA-256 and SHA-3 *466*
13.4.3.2 HMAC *466*
13.5 Applications of Cryptography *467*
13.5.1 Secure Communication *467*
13.5.2 Data Integrity and Authentication *468*
13.5.3 Digital Signatures *468*
13.6 Cryptanalysis and Cryptographic Attacks *469*
13.6.1 Brute Force Attacks *469*
13.6.2 Frequency Analysis *470*
13.6.3 Known Plaintext and Chosen Plaintext Attacks *470*
13.7 Conclusion and Future Directions *471*
13.7.1 Quantum Cryptography *471*
13.7.2 Post-quantum Cryptography *472*
13.8 A Closing Remark on a Nearly Unbreakable Code, Resolved After 50 Years *472*

Index *477*

About the Authors

Alfred Basta, PhD, CMMC-CCA, CISM, CPENT, LPT, OSCP, PMP, CRTO, CHPSE, CRISC, CISA, CGEIT, CASP+, CYSA+, is a professor of mathematics, cryptography, and information security, as well as a professional speaker on internet security, networking, and cryptography. He is a member of many associations, including ISACA, ECE, and the Mathematical Association of America. Dr. Basta's other publications include *Computer Security and Penetration Testing*, *Mathematics for Information Technology, Linux Operations and Administration*, and *Database Security*. In addition, Dr. Basta is the chair of EC-Council's CPENT Scheme Committee. He has worked as a faculty member and curriculum advisor for programming and cybersecurity programs at numerous colleges and universities.

Stephan DeLong has taught mathematics at Tidewater Community College's Beach Campus since 1993. He obtained his Bachelor of Science Degree at Northern Illinois University in DeKalb (Illinois) and his Master of Science Degree at Lehigh University in Bethlehem (Pennsylvania). His PhD work was ended after completing all but his dissertation at Colorado State University. He is a lover of cats, an avid bike rider, tennis player, and surfer, author of several mathematics textbooks, and an aspiring author of noir-styled detective fiction set in the 1970s. He dreams of operating a shuttle boat service from Miami to the Keys and to Bimini on his classic Cigarette racing boat, the *Ultramarine*.

Stavros Basta, CIPP/US, CISM, CPENT, LPTM, CEH, SecurityX, PMP, is a distinguished cybersecurity consultant and accomplished grant writer whose expertise bridges the theoretical foundations of mathematics with practical cybersecurity applications. With extensive experience in the field, Stavros has established himself as a thought leader in the integration of mathematical principles into cybersecurity frameworks. As a trusted advisor, he serves on numerous global advisory boards dedicated to advancing cybersecurity curricula and shaping privacy legislation across international jurisdictions. His multidisciplinary approach to cybersecurity, informed by rigorous mathematical analysis, offers readers a unique perspective on the quantitative underpinnings of modern digital security protocols.

Acknowledgments

I would like to express my profound gratitude to Rebecca Basta for her exceptional contribution to *Mathematics in Cybersecurity*. Her meticulous attention to detail in crafting the end-of-chapter problems and developing the comprehensive test bank has not only enhanced the pedagogical value of this work but has also provided essential resources that will benefit both educators and students alike.

Rebecca's remarkable talent for visual communication is exemplified in the presentations accompanying this book. Her ability to transform complex mathematical concepts in cybersecurity into accessible, engaging visual formats has significantly elevated the instructional potential of these materials, making them invaluable tools for classroom teaching and independent study.

Throughout this scholarly endeavor, Rebecca's unwavering dedication, profound expertise, and steadfast support have been instrumental in bringing this project to fruition. The exceptional quality of her contributions reflects not only her commitment to academic excellence but also her deep understanding of the intricate relationship between mathematics and cybersecurity.

With sincere appreciation and admiration,
Alfred Basta

About the Companion Website

This book is accompanied by a companion website:

www.wiley.com/go/bastamic



This website include:

- PPT & Problem Word File

1

Introduction to Sets

Concepts

- Definition of a set
- Elements of a set
- Cardinality of a set
- Notation (using braces, ellipsis, rosters, set builder notation)
- De Morgan's Laws

> ***Augustus De Morgan*** *(born June 27, 1806, Madura, India—died March 18, 1871, London, England) was an English mathematician and logician whose major contributions to the study of logic include the formulation of De Morgan's laws and work leading to the development of the theory of relations and the rise of modern symbolic, or mathematical, logic.*
>
> Encyclopedia Brittanica

1.1 Introduction

The landscape of history of mathematics, logic, and computer science is populated by gigantic figures: Sir Isaac Newton, Aristotle, Alan Turing, Galileo Galilei, Ludwig Wittgenstein, and countless others. As we embark on a journey to understand mathematics and its relationship to cybersecurity, we remark that not all of what we will discuss here will be straightforward (let alone simple!). Many of the results we present and use will be named after men and women of genius, and we will pause to honor them in our writing; it is to them that we all owe a measure of gratitude, awe, and respect.

We are, metaphorically speaking, walking in the footsteps of giants.

Be confident, however, that you are eminently capable of mastering the material we are about to present. You will emerge from the course with new ideas and techniques, new skills and insights, and a sense of accomplishment that will see you through the challenges that lie ahead.

Definition *Theorem: A theorem, in mathematics and logic, a proposition or statement that is demonstrated. Very loosely, the word can be interpreted as "a rule."*

Definition *Set: A well-defined collection of objects, typically gathered according to some unifying trait, denoted by a capital letter such as A, B, and C.*

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

Definition *Set Theory: A branch of mathematics that deals with the properties of well-defined collections of objects and serves as a basis for precise and adaptable terminology for the definition of complex and sophisticated mathematical concepts.*

1.1.1 The Definition of a Set

A set is a well-defined, unordered collection of objects containing no duplicate members. By "well-defined," we mean that membership in the collection is unambiguous, not open to opinion or interpretation, and strictly capable of determination through an objective investigation of the facts. By "unordered," we specify there is not necessarily an importance to the order in which the members of the set appear within the collection.

The set of "all wealthy countries in the world," for example, is not well-defined, while the set of "countries within the European Union" (EU) is well defined. The perception of what it means to be a wealthy country is open to debate and may not be interpreted the same way by different people. Consequently, whether Senegal is a member of the set of all countries within the European Union can be definitively established, while whether Senegal is a wealthy country cannot (unless we specifically define what we mean by the concept of "wealthy").

The set of numbers {1,2,3,4,5} has an identical list of members as the set of numbers {3,1,4,2,5}. The order of the numbers is not the same, but this does not change the membership of the set.

In our text, sets will primarily be named using capital letters from the English alphabet. There are exceptions to this, as we will see, but this is the convention we will generally adopt. Thus, we can say $S = \{1,2,3,4,5\}$; any presentation of precisely this collection, however ordered, is an instantiation of the set S.

1.1.2 The Elements of a Set

We have mentioned a set is a well-defined, unordered collection of objects containing no duplicate members. The objects within or members of the set are called "elements."

It is common, we should say, for a single concept to be referred to in multiple ways. While occasionally creating a mild confusion, it is something with which we must deal. For example, the English word "dog" and the French word "chien" refer to the same concept, but are in no way similar.

Much of this variety in nomenclature arises from the long, slow climb of human knowledge. We take for granted that information today can be shared rapidly (even instantaneously), but this was not the case throughout history. Often, identical ideas were developed simultaneously in far-flung locations of the globe, with the developers using wildly different notation and terminology.

A particular symbol is used to represent the concept of membership in the set: $\in$. This is read as "is an element of," or some variant of this phrase. A certain amount of care must be taken in the use of this symbol, because subtleties of set theory can result in our misstatement of relationships.

For instance, suppose we consider the earlier mentioned set $S = \{1,2,3,4,5\}$. It would be true to state "1 is an element of the set S," which we represent by $1 \in S$. It would not be true, on the other hand, to state $\{1\} \in S$. This is a subtle but important distinction. The latter statement tells us: "the set consisting of the number 1 is an element of the (larger) set S." This is incorrect, though not apparent. If we had specified the membership of the set S was {{1},2,3,4,5}, THEN it would be correct to state $\{1\} \in S$, but incorrect to state $1 \in S$.

This is a fine point of "attention to detail," wherein we must pay very close attention to (and take great care in) our use of language. This, we point out, is one of the "reasons" all students

are required to take classes in mathematics, as well (in an ideal world) as logic: the science is the ultimate proving ground for attention detail – a skill most of us believe we possess and yet in which we are not proficient.

In any of the precise sciences (mathematics, medicine, computer science, engineering, etc.), great care should be taken in the selection and use of notation and terminology. It is assumed we "know what we're talking about," and we should comport ourselves as such. A trivial example is the casual description of a particular shape as being a "parabola," when it merely *resembles* a parabola. Parabolic curves and shapes have properties unique to that particular categorization of objects and if we identify something as being a parabola, readers will assume they can exploit those properties. The results of such a misinterpretation could be disastrous.

There is a related notation about set membership with which we should be aware: the symbol indicating nonmembership. How are we to indicate a particular item is NOT an element of a given set? There is a convention widely used in logic and mathematics, placement of a "slash" through an already-established symbol, to indicate the concept of "not."

That is, since $\in$ represents "is an element of," the associated symbol $\notin$ represents "is not an element of." Therefore, since we have identified $S = \{1,2,3,4,5\}$, we can say $6 \notin S$.

1.1.3 Cardinality of a Set

Occasionally, it is important to refer to the quantity of elements lying within a set. We refer to this as the *cardinality* of the set.

The notation for cardinality varies; in this text, we will use two notations to indicate the concept: $|S|$ and $n(S)$. Each of these is read as **the cardinality of set *S***, or **the number of elements in the set *S***.

For our set $S = \{1,2,3,4,5\}$, we can say $|S| = 5$, or that $n(S) = 5$.

The cardinality of a set can reach extreme values. By this we mean the low end for the value of cardinality is zero (more will be said of this later), while the high end is infinity (∞). When the cardinality of a set is not infinite, we say it is finite. A specific value, however large, is considered finite. For instance, a set with one trillion elements is a large but finite set.

The cardinality of a finite set is a *discrete* concept. This means there is a definitive number of elements in a finite set. We cannot have, for instance, a set with cardinality π. Since we are determining a count of the members of the set, the result must be a whole number.

1.1.4 The Notation of Sets

The set $S = \{1,2,3,4,5\}$, which presents the membership as a list, incorporates two ideas we should make clear.

First is that the list of elements is surrounded by "curly parentheses," which we call **braces**. Note there is a distinction in mathematics between parentheses, (), brackets, [ ], and braces, {}. We also refer to left and right parentheses/brackets/braces.

Second is the list itself. We refer to a list of this sort as a ***roster*** of membership. It is permissible to informally describe this as a "list of elements," but our formal reference will be as a "roster."

It is, obviously, only convenient to utilize roster notation in the event the set consists of a (relatively) small group of elements. A lengthy roster would be unwieldy, and thus the choice not to use roster notation would be a matter of practicality. The reason we use the term "relatively" in this description is that "small" is a subjective word: what do we each mean by the idea of a "small" group of elements? This can vary from person to person and is often a judgment call. We will use

caution when incorporating relative terms; it is typically too vague within the sciences to be loose with our phrasing, and we shall strive for precision.

We will presume, however, that for now that the idea of a "small group of elements" is reasonably agreeable to all of us, and that it is something about which we will not argue.

There are conventions we can adopt that circumvent the issue of "large" rosters. These are situationally driven and cannot always be used (for reasons we will make apparent through example).

Consider the set $T = \{1,2,3,4,5,6,7,8,9,10,11,12,13\}$. Is this a "large" set? Obviously, this could be debated; however, assume for the moment this set is "large."

In this case, the elements of the set follow a discernible pattern: they are the natural numbers from 1 through 13, inclusive. This might be less evident if the order in which the roster is given were "shuffled." Nevertheless, as presented the membership is rather clear. We can condense the list through the use of **ellipsis**, which is a series of three "dots," intended to indicate the observable pattern continues as seen, beyond the shown portion of the roster.

That is, the set T could be shown as $T = \{1,2,3,\ldots,13\}$.

We can see, based on the first few members displayed, the pattern existing among the elements. The ellipsis indicates the pattern continues through the number 13, where it terminates.

If no value follows the ellipsis, the list does not terminate; it is, instead, infinite. In other words, $V = \{1,2,3,\ldots\}$ is the entire list of natural numbers. Combining the notions of roster notation, ellipsis, and cardinality, we now state:

$S = \{1,2,3,4,5\}$, and $\mathrm{n}(S) = 5$
$T = \{1,2,3,\ldots,13\}$, and $\mathrm{n}(T) = 13$
$V = \{1,2,3,\ldots\}$, and $\mathrm{n}(V) = \infty$

There are other ways in which sets can be presented. A very commonly used (yet often misunderstood) method is **set builder notation**. This method incorporates braces, a separation bar delineating an identification of a symbol for set members and a **defining characteristic** for the elements of the set.

A defining characteristic is a quality or property possessed by the elements of the set. This can be established in written form or using an equation. The format we will adopt for set builder notation will be:

$$\{\textit{general symbol representing an element} \mid \textit{defining characteristic}\}$$

That notation may appear cumbersome; keep in mind it is merely a technical format and is presented in an extremely general form above.

Example 1.1 Describe the membership of the set, using words and a roster.

$$T = \{x \mid x \textit{ is a natural number less than or equal to } 13\}$$

Solution: The notation is read as "T is equal to the set of elements represented by the letter x, such that x is a natural number less than or equal to 13." We can make statement more concise by stating "T is the set of natural numbers less than or equal to 13," or "T is the set of natural numbers less than 14." An expression of set membership using words can always be modified somehow while still accurately describing the members.

A roster of the members of T is $\{1,2,3,4,5,6,7,8,9,10,11,12,13\}$, or $\{1,2,3,\ldots,13\}$.

Example 1.2 Provide a roster of the members of the given set E.

$$E = \{c|c \textit{ is a country in the European Union}\}$$

This would be read as, "E is equal to the set of elements represented by the letter c, such that c is a country in the European Union," Or some close variant of this description. A roster of set E is:

$E = \{$Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden$\}$.

One sees now the utility of set-builder notation: it is much more economical to produce the notation with the description of requirement for set membership rather than creating a roster.

An important observation to make is that the "vertical line" within the braces is always read as "such that."

A final example, more mathematical, is:

$$P = \{(x, y) |\ 3x - 4y = 12\}$$

We read this as, "P is equal to the set of ordered pairs represented by (x,y), such that $3x - 4y = 12$." Recall this criterion indicates the ordered pairs represent an infinitely large set of points all lying on the graph of the equation $3x - 4y = 12$ on the Cartesian Plane. It is worthwhile to remind ourselves that a straight line in the Plane is an example of a "curve." Curves are drawings in the plane (having no width) that contain all the points lying on the graph of an equation. It seems odd, in a way, to refer to a straight line as a "curve," but this is due to the definition offered a moment ago. There is a concept called "curvature" that describes "how curvy" a line is in the Plane, and a straight line is a curve with no curvature.

1.2 Types of Sets

- Empty set (null set)
- Singleton set
- Finite and infinite sets
- Special mathematical sets
- Subset and superset
- Power set

We now turn our attention to specific **types of sets**. When we refer to types of sets, we mean individual sets whose nature or properties elevate them to a standard where they are deserving of specific consideration and nomenclature.

1.2.1 The Empty Set (The Null Set)

In the first section of this chapter, we introduced the notion of cardinality. At that time, we mentioned there were extreme values of cardinality, the first extreme being a set having cardinality zero.

Such a set is called the **empty set**, or the **null set**, represented in one of several ways: first, a defining characteristic that is incapable of being met by any object whatsoever; second, a set of empty braces; third, the empty set symbol: ∅.

> *Note: There are organizations, such as the United States Military, where it is common to use a figure similar to the empty set symbol as a representation of the numeral "0." While this is routinely used, it will create confusion for us here. Therefore, we shall adopt the convention the empty set symbol shall NOT be used synonymously with the numeral 0.*

An example of a defining characteristic impossible of fulfillment would be "the set of all odd numbers that are exactly divisible by 2." No odd numbers are exactly divisible by 2 (this is a consequence

of the definition of "odd numbers," which are "all numbers capable of representation in the form (2k + 1)," where k is an integer), and therefore the set of objects being described is empty. Another illustration is "the set of all human beings currently alive who were born in the 19th century." Since no such human beings exist, this set is empty.

However presented, the cardinality of the empty set is understood to be zero.

Recalling our earlier comment that there exist peculiarities and nuances within the theory of sets, we point out the following issues:

1) { } or ∅ are both used to represent the empty set, or null set
2) The set given by {∅} is not the same as the empty set, or null set

The first point highlights the different notations for the empty set; the second is a subtlety of detail. When we write $N = \{\emptyset\}$, we are referring to a set with one element … that element is the empty set. The cardinality of this set is 1; we will revisit this illustration later in this section, when we introduce the notion of a power set.

The second issue emphasizes a common error we must avoid: imprecise use (or lack of understanding) of notation. We have said ∅ is the empty set; the set {∅} is a set having as a member the empty set. While this may seem bewildering to you, consider the following illustration: suppose we use a thought experiment in which you have an empty cube of glass…the cube can be imagined to be the empty set. Next, suppose you have a cloth bag into which you place the cube. Though the cube remains the empty set, it is now contained within the bag; that is, the bag contains the empty set. Do you see how the two situations are different?

1.2.2 Singleton Sets

Sets having cardinality 1 are referred to as singleton sets. Examples are the sets $S = \{a\}$, $P = \{14\}$, $N = \{\emptyset\}$, $T = \{x | x \text{ is a natural number equal to its own square}\}$, and $V =$ the set of all states in the United States whose names begin with the letter *R*.

If we wish to maintain full generality, we could create a "generic" singleton set, $\{x\}$, with no specification as to the nature of the single element.

Though it may seem frivolous to concern ourselves with this terminology, singleton sets are fundamental components of set theory. The singleton sets are useful for expressing qualities of uniqueness and serve as building blocks for more elaborate sets.

1.2.3 Finite and Infinite Sets

The cardinality of a set may be finite (i.e., membership in the set is limited to a particular, possibly very large, quantity of elements) or infinite (there are an unlimited number of elements in the set). Essentially, any specific numerical value you can imagine is finite. At the moment of this writing, the National Debt of the United States of America is $34,687,005,429,763. Though extremely large, this number is finite.

It is worth noting the immensity of this value! Read aloud, this is "thirty-four trillion, six hundred eighty-seven billion, five million, four hundred twenty-nine thousand, seven hundred sixty-three dollars." Were we to count from 1 to this number, allowing ourselves no rest and taking just one second to read each number, it would take us over one million years to do so!

Care should be taken when referring to and using infinity within statements. We should not assert, for example, that space is infinite. This has not been proven, so we cannot make this claim. Moreover, intuition tends to break down when contemplating infinite quantities.

As an illustration, suppose we were told to find the sum of infinitely many positive numbers. The natural conclusion one would draw is that the sum is infinity. However, an elementary example proves this not to be the case! Consider the progression of numbers 1, 1/2, 1/4, 1/8, 1/16, ... We cannot dispute this list of numbers contains only positive values, and the pattern is clear (the denominator doubles from one number to the next). However, the values (when added) fail to reach 2.

The point of the last example serves to emphasize the bizarre nature of infinity, and the failure of intuition when considering the concept. This will not typically be of concern to us in this text, but suggests it is worthy of substantial investigation. The earliest significant work on the topic was performed by Georg Ferdinand Ludwig Philipp Cantor (1845–1918), a Russian-born mathematician who played a pivotal role in the creation of the field of mathematics we call Set Theory.

1.2.4 Special Mathematical Sets

There are collections of numbers that are used so frequently that it becomes convenient to identify them using special characters. Here we will list several, acknowledging that more do exist.

$\mathbb{N}$: the set of natural (counting) numbers

$\mathbb{W}$: the set of whole numbers

$\mathbb{Z}$: the set of integers

$\mathbb{Q}$: the set of rational numbers

$\mathbb{R}$: the set of real numbers

These sets are considered familiar to us, and we will not express detailed definitions here.

1.2.5 Subset and Superset

With the notion of a set now well in hand, we can begin to explore relationships between sets. The first such concept will be the "subset" and "superset" relationships.

Set A is a subset of set B if every member of set A is also a member of set B. The notation we use to indicate this is either $A \subset B$ or $A \subseteq B$, though those are not identical relationships. The first refers to a particular type of subset, called a PROPER subset. The second is a more inclusive and general subset notation.

Set A is a proper subset of set B if all the elements of set A are elements of set B, with the understanding that B contains additional elements not also contained in set A. For instance, the set $\mathbb{Q} \subset \mathbb{R}$. We could more generally say $\mathbb{Q} \subseteq \mathbb{R}$ and still be correct, but we will adopt the convention that we will use the proper subset notation if it is clear a particular subset is a proper subset of another set.

Note, once again, the peculiar nature of infinite cardinality. Though it is true $\mathbb{Q} \subset \mathbb{R}$, it turns out each has infinite cardinality. This appears to "fly in the face" of the assertion that $\mathbb{R}$ contains elements not in $\mathbb{Q}$, and suggests there may be "degrees of infinity," or some similar concept. As mentioned earlier, we will not explore that notion further, though it is definitely worthy of study.

The superset relationship is the opposing concept for subsets. That is, if set A is a subset of set B, then set B is considered a superset of set A. Therefore, $\mathbb{R}$ is a superset of $\mathbb{Q}$. We may express this by reversing the notation for subsets, using either $\supset$ or $\supseteq$. In other words, $\mathbb{R} \supset \mathbb{Q}$.

As an additional remark, we note the empty set is considered to be a subset of every other set, and that a given set is a subset of itself. The latter point seems odd to many who first encounter it, and thus we say set A is an improper subset of itself (yet, still a subset!).

1.2.6 Power Set

We have presented the concept of a subset of a given set and now may contemplate the complete list of subsets of that given set. The set of all subsets (including the empty set and the improper subset) of set A is referred to as the POWER SET of set A and is symbolized as $\mathcal{P}(A)$ or as 2^A.

Observe $\mathcal{P}(\emptyset) = \{\emptyset\}$, because the empty set is a subset of every set, including itself. Moreover, $\mathcal{P}(\{x\}) = \{\emptyset, \{x\}\}$, and $\mathcal{P}(\{x, y\}) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$.

Though we could continue this list of illustrations, we make the following observation: if $n(A) = a$, then the cardinality of $\mathcal{P}(A) = n(\mathcal{P}(A)) = 2^a$. Adding an element to a set doubles the number of subsets that set possesses.

Listing full membership of the power set of a set can be extremely tedious. Owing to the rapidity of exponential growth, a set with (e.g.) 10 elements will have $2^{10} = 1024$ subsets!

Despite its unwieldiness, the power set is fundamental in set theory and has applications in various areas of mathematics and computer science, particularly in combinatorics (the branch of mathematics dealing with combinations of objects belonging to a finite set in accordance with particular constraints) and logic.

1.3 Operations on Sets

- Union of sets
- Intersection of sets
- Cardinality of unions and intersections
- Difference of sets

1.3.1 Union of Sets

When two or more sets are combined, we say that we are creating the "**union**" of the sets. We use a symbol resembling a capital "*U*" in the notation for union, though it is typically shorter and wider than the letter to which it bears a resemblance.

If the sets are A and B, the union of those sets is denoted $A \cup B$. By construction, both A and B are subsets of their union, and their union is hence a superset of both A and B. We refer to the separate sets A and B as the "component sets" of the union. If one set is a subset of the other, the union is equal to the superset.

Keep in mind the roster of a set does not contain duplication of elements, and so if there are elements common to both sets, they should be listed only once in the union.

Consider the set of students in your mathematics class and the set of members of your family; the union of the two sets would be the set of all the members of your math class, together with the members of your family.

We cannot say, at this time, the cardinality of the union is the sum of the cardinalities of the component sets. Why this is the case should, with some thought, be clear to you. We will address this matter later in the section, but it is worthwhile to see if you recognize the reason for this now.

Example 1.3 Let $A = \{1,3,5,7,9\}$ and $B = \{2,3,4,5,6\}$. What is $A \cup B$?

Solution: Combining the two sets, and noticing the elements 3 and 5 appear on both rosters, we find $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 9\}$.

Referring to our remark from just prior to the example, observe $n(A) = 5$ and $n(B) = 5$, but $n(A \cup B) \neq 5 + 5 = 10$. Now do you see why the cardinality of the union might not equal the sum of the component set cardinalities?

Example 1.4 Let $A = \{x \in \mathbb{R} \mid x^2 = x\}$ and $B = \{x \in \mathbb{R} \mid x^2 - x = 6\}$. What is $A \cup B$? Give your answer as a roster.

Solution: One way to give the union would be to employ the word "or" into our set-builder notation; that is, $A \cup B = \{x \in \mathbb{R} \mid x^2 = x \text{ or } x^2 - x = 6\}$. This, however, is not a roster (as was requested) and is uninformative. We know next to nothing about the membership of the set if we employ this technique.

We have several methods of attack open to us, and we shall explore both (we may discover something interesting). For set A, what is the roster of its members? We know the criteria for membership, so we solve the equation mentioned in the property area of set-builder notation.

$$x^2 = x$$
$$x^2 - x = 0$$
$$x(x-1) = 0$$
$$x = 0 \quad \text{or} \quad x - 1 = 0$$
$$x = 0 \quad \text{or} \quad x = 1$$

Therefore, A = {0, 1}.

For set B, we must solve:

$$x^2 - x = 6$$
$$x^2 - x - 6 = 0$$
$$(x-3)(x+2) = 0$$
$$x - 3 = 0 \quad \text{or} \quad x + 2 = 0$$
$$x = 3 \quad \text{or} \quad x = -2$$

Therefore, $B = \{3, -2\}$.

Combining our results, $A \cup B = \{0, 1, 3, -2\}$.

Now consider the following (flawed) reasoning: we know the members of the union must satisfy $x^2 = x$ or they must satisfy $x^2 - x = 6$. By substitution, since $x^2 = x$, we must have $x^2 - x^2 = 6$, or $0 = 6$. This is a contradiction, so the union of the two sets must be empty.

What we have done with this second approach, while appearing completely reasonable, was to change the nature of the problem. By using substitution in the manner described, we have actually resolved the issue regarding which values of x are such that $x^2 = x$ AND $x^2 - x = 6$. This subtle change in language has led us completely astray and points out to us that we must be very cautious in our choice of wording, and of our application of mathematical techniques!

1.3.2 Intersection of Sets

When the common elements of two or more sets are considered and gathered into a set, we say we are creating the "**intersection**" of the sets. We use the upside-down form of the union symbol to represent intersection.

If the sets are A and B, the intersection of those sets is denoted $A \cap B$. By construction, $A \cap B$ is a subset of both A and B, and both A and B are supersets of their intersection. We again refer to the separate sets A and B as the "component sets" of the intersection.

If A and B have the same set of elements, then the intersection is all elements in either of the component sets. If A and B have no common elements, the intersection is the empty set.

Example 1.5 Let A = the set of even integers between 0 and 20, inclusive, and B = the set of integers divisible by 3 between 0 and 20, inclusive. Find $A \cap B$.

Solution: Set $A = \{0,2,4,6,8,10,12,14,16,18,20\}$, set $B = \{0,3,6,9,12,15,18\}$. The members common to the two sets are 0,6,12, and 18, so $A \cap B = \{0,6,12,18\}$.

Example 1.6 Let $A = \{2,4,6,8,10\}$, $B = \{x | \, x^2 - 2x - 3 = 0\}$. Find $A \cap B$.

Solution: The membership of A is explicit, but the membership of B is not so clear. We will work to find the membership of B as a roster, and then the intersection should be apparent.

We begin by solving $x^2 - 2x - 3 = 0$. The expression on the left-hand side of the equation factors as $(x-3)(x+1) = 0$, and the zero-product rule of algebra tells us we must have either $x - 3 = 0$ or $x + 1 = 0$, which yields $x = 3$ or $x = -1$, so $B = \{3,-1\}$.

Next, observe there are no members common to A and B, and so we conclude the intersection is the empty set, and write $A \cap B = \emptyset$.

1.3.3 Cardinality of Unions and Intersections

Consider the following paraphrase of the definition of the union of two sets: "the union of two sets is created by combining the two sets together, without including duplications of common elements."

With this in mind, we can reason out the cardinality of the union. The cardinality of the union will be the sum of the cardinalities of the component sets, subtracting off the number of duplications of elements. The duplication of elements is precisely the intersection of the two sets, those elements common to both sets.

That is,

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Rearrangement using algebra gives us a formula for the cardinality of the intersection:

$$n(A \cap B) = n(A) + n(B) - n(A \cup B)$$

Example 1.7 We saw that for A = the set of even integers between 0 and 20, inclusive, and B = the set of integers divisible by 3 between 0 and 20, inclusive, the intersection was $\{0,6,12,18\}$. Show the formula for the cardinality of the intersection and union holds for these sets.

Solution: We find $n(A) = 11$, $n(\mathrm{B}) = 7$, and $n(A \cap B) = 4$. Therefore, $n(A \cup B) = 11 + 7 - 4 = 14$. Further, we find $n(A \cap B) = 11 + 7 - 14 = 4$.

1.3.4 Difference of Sets

Recall that $A \subset B$ and $A \subseteq B$ indicate B is a superset of A. This means (in the case of a proper subset relationship) there exist elements in set B that are not in set A. This list of additional elements is called the "difference between A and B," denoted by $B - A$. Observe the superset is listed first.

The difference between sets could be infinite, if the superset is infinite. If both sets are finite, the difference is finite, and if the subset A is not a proper subset of B, the difference could be empty.

Note the rationale for using the term "empty set" becomes more evident as we proceed. In the case of the difference of sets, imagining the "additional elements" are in a theoretical "bag," if there is no difference between the sets then the bag is empty.

Note further this concept of "no difference" suggests, as we will soon discover, there exists a notion of set equality.

Example 1.8 For the sets $A = \{3,4,5,6,7,8\}$ and $B = \{1,2,3,4,5,6,7,8\}$, what is set $B - A$?

Solution: Set B contains elements 1 and 2 that are not in set A, so $B - A = \{1,2\}$.

Example 1.9 Let $A = \{x \mid x$ is an even non-negative integer less than 10$\}$, $B = \{x \mid x$ is a whole number less than 10$\}$. What is $B - A$?

Solution: The roster of set A is $\{0,2,4,6,8\}$, set B is $\{0,1,2,3,4,5,6,7,8,9\}$. We see that B is a superset of A, and that $B - A = \{1,3,5,7,9\}$.

Example 1.10 Let $A =$ The set of integers whose absolute value is less than or equal to 10, and $B = \{0,1,2,3,4,5,6,7,8,9,10\}$. What is $B - A$?

Solution: The sets A and B coincide; that is, A and B have the same membership. Therefore, $B - A = \emptyset$.

1.4 Set Relations

- Equality of sets
- Equivalence of sets
- Disjoint sets
- Universal set
- Complement of a set
- Venn diagrams

Section 1.3 closed with the notion of the "difference between sets." If there is no difference (i.e., if the difference of sets is empty), the sets are equal.

Equal sets possess precisely the same elements as one another. The elements may be described differently, have their order shuffled, or presented in different manners, but the list of elements in the sets is precisely the same. This means equal sets have precisely the same cardinality, though sets with the same cardinality need not be equal.

Consider the following sets:

$A = \{x \mid x$ is an even natural number less than or equal to 100$\}$
$B = \{2,4,6,\ldots,100\}$
$C = \{x \mid x = 2n, n \in \mathbb{N}, 1 \leq n \leq 50\}$

The sets A, B, and C are all equal. We have seen here and in the past that sets can be represented in multiple ways, and it is the membership (not the description) that establishes their equality.

The relationship of equality between sets possesses three fundamental properties we now point out explicitly:

1) Sets should be equal to themselves, that is, it should always be true that $A = A$. This requirement is referred to as *reflexivity*.
2) If $A = B$, then $B = A$. This requirement is referred to as *symmetry*.
3) If $A = B$ and $B = C$, then $A = C$. This is referred to *transitivity*.

These properties hold for many other relationships within other sets, including the real numbers. In particular, we shall see the properties hold for "directed graphs," which we shall encounter later in this work.

1.4.1 Equivalence of Sets

Informally, we can say that sets having precisely the same cardinality are said to be **equivalent** sets. It is evident that equal sets are also equivalent, but that the converse is not necessarily true.

That is, $A = \{1,3,5,7\}$ and $B = \{\alpha, \beta, \gamma, \delta\}$ both have cardinality 4, and so are equivalent sets. By inspection, we see the sets are not equal, as their elements are not identical.

The notation $A \sim B$ shall indicate the sets A and B are equivalent.

Definition Let be sets. We say that A is equivalent to B if and only if there exists a bijection $f : A \to B$. In the case of the example above, we can define such a bijection as:

$$f(1) = \alpha, f(2) = \beta, f(3) = \gamma, f(4) = \delta$$

Clearly, this function is both one-to-one and onto and serves as such a bijection (there could be many bijections between two equivalent sets; finding any particular one suffices).

In the matter of set equivalence, (as was the case in the concept of set equality) the reflexive, symmetric, and transitive relationships exist:

1) Things should be equivalent to themselves. That is, it should always be true that $A \sim A$.
2) If $A \sim B$, then $B \sim A$.
3) If $A \sim B$ and $B \sim C$, then you want $A \sim C$.

Note: These three qualities are, in general, referred to as the properties of an ***equivalence relation***.

1.4.2 Equivalence and Non-equivalence of Infinite Sets

We recall once more our admonition that one must be cautious when dealing with the concept of infinity. Are sets with infinite cardinality equivalent to one another? This is a challenging question, the answer for which can vary depending upon the situation: some infinite sets can be shown to be equivalent while others cannot.

The explanation for this once more falls upon our loss of intuition when it comes to infinity. Clearly, $\mathbb{N} \subset \mathbb{Q}$, because there are rational numbers that are not natural numbers. This situation presents us with a conundrum we will not explore in this course, though we do encourage you to pursue this issue as time permits.

Example 1.11 Show the sets $\mathbb{N}$ and $\mathbb{W}$ are equivalent.

Solution: The sets are each infinite, and we can readily produce a bijection between them. Consider the function $f(x) = x - 1$ from $\mathbb{N}$ to $\mathbb{W}$. The function is clearly one-to-one and onto, and thus serves as a sample bijection. Therefore, the two sets are equivalent.

Example 1.12 Show the sets $\mathbb{N}$ and $\mathcal{P}(\mathbb{N})$ are not equivalent.

Solution: Since the natural numbers form an infinite set, the power set of the natural numbers is also an infinite set. However, we now show the sets are not equivalent.

There are many one-to-one functions from $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$, such as $f(n) = \{n\}$. The image, obviously, is an element of $\mathcal{P}(\mathbb{N})$. Notice the range of f, however, is a subset of $\mathcal{P}(\mathbb{N})$, but not equal to the entire set $\mathcal{P}(\mathbb{N})$.

Suppose the two sets ARE equivalent. Then there exists a bijection f that maps $\mathbb{N}$ to $\mathcal{P}(\mathbb{N})$. That is, every natural number is associated in a one-to-one, onto relationship with some set within the power set. Let us define a new set: $E = \{x \in \mathbb{N} \mid x \text{ is not an element of } f(x)\}$. The set E "collects"

exactly those elements of S that are not in the subset they supposedly got paired with under the function f. Because E is a subset of $\mathbb{N}$, D is a member of $\mathcal{P}(\mathbb{N})$, which implies there is an element e in $\mathbb{N}$ such that $f(e) = E$. Consider the non-obvious question: is "e" in the set E, or not? If it is, then by the definition of set E, e is not an element of $f(e) = E$, which is a contradiction. On the other hand, if e is not an element of the set E, then by definition e is an element of $f(e) = E$, which is again a contradiction. Since we were supposing a bijection did exist and that led to contradictions, no such bijection exists.

1.4.3 Disjoint Sets

We have mentioned there is the possibility the intersection of two sets is empty. That is, two given component sets could have no members in common. When this happens, we say the sets are "disjoint."

The set of all even integers is disjoint from the set all odd integers, for instance. The set of human beings currently alive and the set of human beings who fought in The Revolutionary War are disjoint sets.

Note the consequence for the formulas, for cardinality of intersections and unions of sets, should the component sets be disjoint:

For unions:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$
$$n(A \cup B) = n(A) + n(B) - 0$$
$$n(A \cup B) = n(A) + n(B)$$

For intersections:

$$n(A \cap B) = n(A) + n(B) - n(A \cup B)$$
$$0 = n(A) + n(B) - n(A \cup B)$$
$$n(A \cup B) = n(A) + n(B)$$

As one would expect, the two formulas simplify to the same final expression. This is a good thing; had they not, there would be a fundamental problem with our notions of cardinality and our newly introduced quality of sets being disjoint.

This reminds us of an important and fundamental fact: mathematical ideas should not conflict with one another. If it appears they do, then there is either a problem with calculations or with definitions. Whatever the case, such conflicts must be investigated and resolved.

1.4.4 Universal Sets

When considering sets, an early step is to establish a context within which the set exists. It is customary to define this context by means of an all-encompassing set called the universal set, which we shall symbolize by U.

In each situation, all sets considered within a particular problem must take their members from that universal set. An example of a universal set could be a particular set of numbers, or a particular group of individuals. If no universal set is specified, it may be that its composition is irrelevant for the purposes of the problem, and so we can take it to be some larger set within which the members of our sets reside.

For a specific set, the universal set could vary from example to example. Suppose that our set was the set of all persons currently enrolled at your college. The universal set could (unless specifically stated) be the set of all persons within the town in which the college lies, the set of all persons

enrolled at any college at all within the United States of America, or the individuals currently located within the United States of America. When the universal set is not made explicit by context, and when the identity of the universal set is of importance, we will state it specifically.

Once the universal set has been identified, we will only consider members from that set for the duration of the problem under consideration. If, for instance, we define U = {3,4,5,6,7,8}, then for the remainder of the problem, no other objects exist. Thus, using that universal set, if we then wanted to create a set consisting of all even numbers from the universal set, and call that set E, we would say E = {4,6,8}. While there are other even numbers from the broader context of all possible numbers, from the perspective of our universal set, these are the only even numbers in existence.

Consider now the letters q and Q. If the universal set were the 26 letters of the English alphabet, then we could rightly claim that both symbols represent the same thing, and thus are the same. If, on the other hand, the universal set is the 52-member collection of upper- and lower-case letters of the English alphabet, then the elements q and Q are distinct, and should not be treated as being one and the same.

The moral to that last example is that we must be very careful when we consider the members of sets because it is possible that the specification of the universal set could affect our observations and analysis.

1.4.5 The Complement of a Set

Given a universal set U, and a set A within that universal set, it is meaningful to speak of the "complement" of the set. Observe this should not be mistaken for the homonym "compliment." The "compliment" of a set might be, "You're a very attractive set." This is not the same as the "complement" of a set!

The complement of set A, represented by A^c or A', refers to the set of elements in the universal set not contained within set A. Without knowledge of the nature of the universal set, it is impossible to accurately describe the complement of set A.

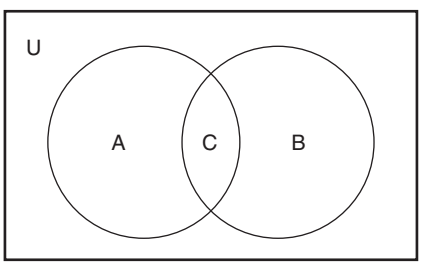
If we let $U = \{a,b,c,d,e,f,g\}$, then identify $V = \{a,e\}$, we find $V^c = \{b,c,d,f,g\}$.

The sets V and V^c are, by definition, disjoint, and their union is equal to the universal set.

1.4.6 Venn Diagrams

A Venn Diagram (named after its inventor, John Venn [1834–1923]) is a useful visual tool to represent sets and their relationships to one another.

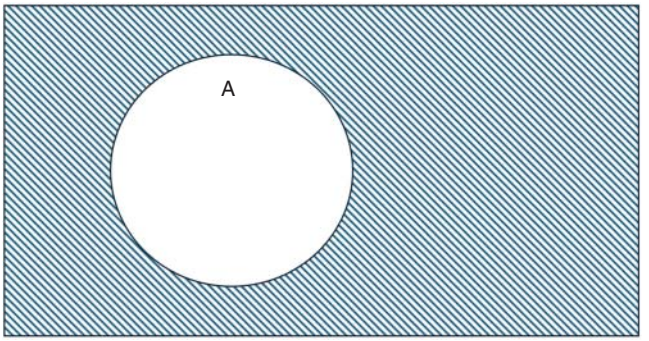
A typical Venn Diagram would look something like this:



In this situation, the two overlapping circles A and B each represent sets, and the area that is common to both circles is labeled C, which is the overlap, or intersection of sets A and B.

The sets are surrounded by a rectangle that encloses them; the rectangle represents what we have identified as the universal set.

When we consider a set drawn in a Venn Diagram, such as A above, we can envision the circle representing A breaking the universal set into two pieces: those things that are in set A, and those things that are not. Thus, the region within the rectangle but outside the circle represents A^c.



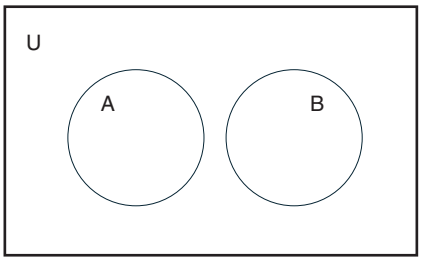
In the image, the shaded area of the rectangular representation of the universal set is A^c.

Consider the following example. Suppose that the universe consists of all the states in the United States. Set A represents the set of states whose names begin with the letter "N." Set $A = \{$Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota$\}$. Those are the states sitting inside circle A. The other 42 states are OUTSIDE of the circle A and comprise A^c.

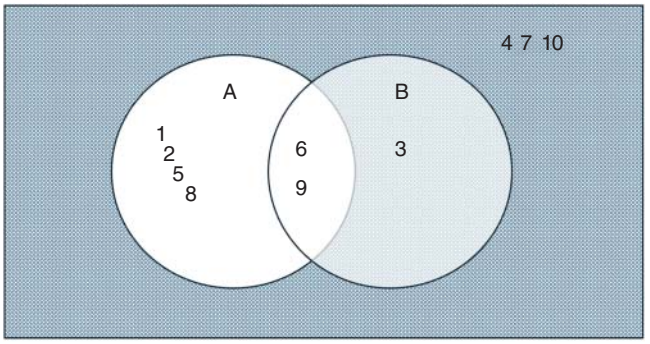
Suppose that set B represents the states in the "contiguous United States." That is, the states that are in the "main part" of the United States, all the connected states.

The area labeled C on diagram would be the "overlap," or "intersection" of the two sets. Since the only two states not within the contiguous United States are Alaska and Hawaii, the region represented by C is empty.

Though, in generality and if it is made clear which elements are within sets A and B, it is permissible to represent this situation with overlapping circles, the disjoint relationship of sets A and B is typically made clear by showing the sets A and B as non-intersecting circles.



Example 1.13 Construct a Venn diagram representation of the universal set $U = \{x \in \mathbb{Z} \mid 0 < x \leq 10\}$, with sets $A = \{1,2,5,6,8,9\}$ and $B = \{3,6,9\}$.



Note the universal set is not labeled with "U." This can be done, but we understand in this text, the all-encompassing rectangle will be the universal set and will suppress labeling of such.

2

Logic

2.1 Introduction to Logic

Logic is a fundamental branch of mathematics that deals with the study of reasoning, inference, and the principles of valid arguments. It provides a framework for analyzing and evaluating the correctness of assertions, statements, and proofs.

In this section, we explore the definition of logic, its importance in discrete mathematics, and the basic building blocks of logical reasoning: propositions and logical connectives.

2.1.1 Definition of Logic

Logic is the study of the principles and methods of reasoning, particularly in the context of mathematical and philosophical arguments. Arguments are a set of statements (two or more premises and a conclusion), and one of the objectives of logic is to determine if a particular argument is valid—we distinguish between the notions of validity and soundness, which we will discuss shortly.

The premises purportedly provide evidence, reasons, and grounds for the conclusion, which is the notion for which the argument is being made. An argument alleges some reasoned connection between the premises and the conclusion, and such a connection is the subject of logical analysis.

The study of logic involves the systematic analysis of the form and structure of arguments, rather than the content of those arguments. The validity of arguments is based on the correctness of their logical form, independent of the truth or falsity of their premises. A premise is a proposition antecedently supposed in an argument; for the purpose of an individual problem in logic, we shall accept each premise as true.

For instance, we might say: ***All pigs can fly. Arnold is a pig. Therefore, Arnold can fly.*** An immediate objection one might raise is: "Pigs can't fly!" In the examination of an argument, we are ignoring this (obviously absurd) assertion's truth or falsity.

Logic is divided into two main branches:

1) Formal logic deals with the abstract structure of arguments and the rules for drawing valid conclusions from given premises. It includes propositional logic, predicate logic, and other formal systems.
2) Informal logic focuses on the analysis of arguments in natural language, taking into account the context, meaning, and practical implications of the reasoning process.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

2.1.2 Importance of Logic in Discrete Mathematics

In mathematics, logic provides the foundation for various concepts and techniques used in the field. In the particular instance of discrete mathematics, the study of mathematical structures that are countable or otherwise distinct and separable, some of the key reasons why logic is important include:

1) Reasoning and proofs: Logic is essential for constructing and validating mathematical proofs. It is used in establishing the correctness of theorems, lemmas, and other mathematical statements by ensuring that the reasoning process follows valid inference rules.
2) Algorithm design and analysis: Logical reasoning is fundamental to the design and analysis of algorithms and procedures for solving a mathematical problem in a finite number of steps that frequently involve recursive operations. It is used to express the preconditions, postconditions, and invariants of algorithms, as well as to prove their correctness and analyze their complexity.
3) Digital systems and computer science: Logic is at the core of digital systems and computer science. It is used to design and analyze digital circuits, develop Boolean algebra, and create logical expressions that govern the behavior of computer programs.
4) Set theory and combinatorics: Logic is closely related to set theory and combinatorics. It helps in defining sets, expressing set operations, and analyzing the relationships between sets. Logical connectives and quantifiers are used extensively in these areas.
5) Decision-making and problem-solving: Logic provides a systematic approach to decision-making and problem-solving. It helps in breaking down complex problems into smaller, manageable components and applying logical reasoning to arrive at valid conclusions.

2.1.3 Propositions and Logical Connectives

The basic building blocks of logical reasoning are propositions and logical connectives. A proposition is a declarative statement that is either true or false, but not both. Propositions are typically represented by lowercase letters (e.g., p, q, and r) and can be combined using logical connectives to form compound propositions.

Definition Logical connectives are logical constants. Connectives can be used to connect declarative statements.

The most common logical connectives are:

1) Negation (NOT): Denoted by the symbol $\neg$, it reverses the truth value of a proposition. If p is true, then $\neg$p is false, and vice versa.
2) Conjunction (AND): Denoted by the symbol $\wedge$, it combines two propositions and is true only when both propositions are true.
3) Disjunction (OR): Denoted by the symbol $\vee$, it combines two propositions and is true when at least one of the propositions is true.
4) Implication (IF–THEN): Denoted by the symbol $\rightarrow$, it represents a conditional statement. If $p \rightarrow q$, then q is true whenever p is true.
5) Biconditional (IF AND ONLY IF): Denoted by the symbol $\leftrightarrow$, it represents an equivalence between two propositions. If $p \leftrightarrow q$, then p and q have the same truth value.

These connectives allow us to express complex logical statements and analyze their truth values based on the truth values of the individual propositions.

2.2 Propositional Logic

Propositional logic is a fundamental branch of logic that deals with propositions and their relationships. It provides a framework for analyzing the logical structure of statements and arguments using logical connectives to combine and manipulate propositions.

2.2.1 Propositions and Compound Propositions

A proposition is a declarative statement that is either true or false, but not both. Propositions are the building blocks of propositional logic and are usually represented by lowercase letters, such as p, q, or r.

Examples of propositions:

- p: "The Earth is round."
- q: "$2 + 2 = 4$."
- r: "Paris is the capital of Italy."

Each of these statements is either true or false. For instance, the propositions "The Earth is round" (p) and "$2 + 2 = 4$" (q) are true, while the proposition "Paris is the capital of Italy" (r) is false.

In logic, a statement's **truth value** indicates whether the claim being made represents the way the world actually is. Some statements do, while others do not. So, any statement in logic can have one of two truth values at any given time: **true** or **false**. A statement must be either true or false; it cannot be neither and cannot be both true and false at the same time.

Compound propositions are formed by combining two or more propositions using logical connectives. The truth value of a compound proposition depends on the truth values of its constituent propositions and the logical connectives used.

Examples of compound propositions:

1) Negation (NOT):
   - If p is the proposition "The sky is blue," then ¬p represents the proposition "The sky is not blue."
   - If q is the proposition "$2 + 2 = 5$," then ¬q represents the proposition "It is not the case that $2 + 2 = 5$."

We pause for a moment to suggest that negations can be poorly stated if care is not taken in the analysis of a proposition. For example, if we attempt to negate the proposition, "Scott gained weight," it would be incorrect to state, "Scott lost weight." Brief consideration tells us that we are committing an assumptive overreach by suggesting that this second statement is the negation of the first. Another possibility exists: perhaps Scott's weight was unchanged? The overreach we have made is the incorrect view that if a person has not gained weight, then they must have lost weight.

Because of this potential dilemma, it is often best (though it may sound stilted) to negate a statement by using the phrase "it is not the case that." For the example involving Scott, the negation of the statement "Scott gained weight" would be correctly negated by stating "It is not the case that Scott gained weight." This leaves open the possibility he either lost weight or that his weight remained the same.

2) Conjunction (AND):
   - If p is the proposition "I like pizza," and q is the proposition "I like pasta," then $p \wedge q$ represents the proposition "I like pizza and I like pasta."
   - If r is the proposition "$x > 0$," and s is the proposition "$x < 10$," then $r \wedge s$ represents the proposition "x is greater than 0 and less than 10."

3) Disjunction (OR):
   - If p is the proposition "I will go to the beach," and q is the proposition "I will go to the mountains," then $p \vee q$ represents the proposition "I will go to the beach or I will go to the mountains."
   - If r is the proposition "The number is even," and s is the proposition "The number is divisible by 3," then $r \vee s$ represents the proposition "The number is even or divisible by 3."

We are describing "or" in the inclusive sense as a logical connective. That is, the statement $p \vee q$ is considered true if p is true, q is true, or both are true. There is another sense of the word "or," called the "exclusive or." If we use "or" in this sense, we mean that either p is true or q is true, but not both. Unless we specifically mention otherwise, we will be using the inclusive or.

4) Implication (IF-THEN):
   - If p is the proposition "It is raining," and q is the proposition "The ground is wet," then $p \rightarrow q$ represents the proposition "If it is raining, then the ground is wet."
   - If r is the proposition "$x > 5$," and s is the proposition "$x > 0$," then $r \rightarrow s$ represents the proposition "If x is greater than 5, then x is greater than 0."

The statement following "if" and preceding "then" is called the antecedent. The statement following "then" is called the consequent.

5) Biconditional (IF AND ONLY IF):
   - If p is the proposition "A shape is a square," and q is the proposition "A shape has four equal sides and four right angles," then $p \leftrightarrow q$ represents the proposition "A shape is a square if and only if it has four equal sides and four right angles."
   - If r is the proposition "x is even," and s is the proposition "x is divisible by 2," then $r \leftrightarrow s$ represents the proposition "x is even if and only if x is divisible by 2."

2.2.2 Logical Connectives and Truth Values

The truth value of a compound proposition depends on the truth values of its constituent propositions and the logical connectives used.

2.2.2.1 Negation (NOT)

Negation, denoted by the symbol ¬, reverses the truth value of a proposition. If p is a proposition, then ¬p (read as "not p") is true when p is false, and false when p is true.

Example 2.1

- Let p be the proposition "The car is red," and assume p is true.
- Then, ¬p represents the proposition "The car is not red." Since we are assuming p to be true, ¬p is false.

2.2.2.2 Conjunction (AND)

Conjunction, denoted by the symbol ∧, combines two propositions and is true only when both propositions are true. If p and q are propositions, then $p \wedge q$ (read as "p and q") is true only if both p and q are true.

Example 2.2

- Let p be the proposition "I have a dog," and q be the proposition "I have a cat."
- Then, $p \wedge q$ represents the proposition "I have a dog and I have a cat."
- $p \wedge q$ is true only if both p and q are true (I have both a dog and a cat).
- If either p or q or both are false (I do not have a dog, or I do not have a cat, or I have neither), then $p \wedge q$ is false.

2.2.2.3 Disjunction (OR)

Disjunction, denoted by the symbol $\vee$, combines two propositions and is true when at least one of the propositions is true. If p and q are propositions, then $p \vee q$ (read as "p or q") is true if either p or q or both are true.

Example 2.3

- Let p be the proposition "I will watch a movie," and q be the proposition "I will read a book."
- Then, $p \vee q$ represents the proposition "I will watch a movie or I will read a book."
- $p \vee q$ is true if either p is true (I will watch a movie), q is true (I will read a book), or both p and q are true (I will do both).
- $p \vee q$ is false only if both p and q are false (I will neither watch a movie nor read a book).

2.2.2.4 Implication (IF–THEN)

Implication, denoted by the symbol $\rightarrow$, represents a conditional statement. If p and q are propositions, then $p \rightarrow q$ (read as "if p then q") is false only when p is true and q is false; in all other cases, the implication is true.

Example 2.4

- Let p be the proposition "You study hard," and q be the proposition "You will pass the exam."
- Then, $p \rightarrow q$ represents the proposition "If you study hard, then you will pass the exam."
- $p \rightarrow q$ is true in the following cases:
- If p is false (you do not study hard), regardless of whether q is true or false.
- If p is true (you study hard) and q is true (you pass the exam).
- $p \rightarrow q$ is false only when p is true (you study hard) and q is false (you do not pass the exam).

2.2.2.5 Biconditional (IF AND ONLY IF)

Biconditional, denoted by the symbol $\leftrightarrow$, represents an equivalence between two propositions. If p and q are propositions, then $p \leftrightarrow q$ (read as "p if and only if q") is true when p and q have the same truth value, that is, both are true or both are false.

Example 2.5

- Let p be the proposition "A number is even," and q be the proposition "A number is divisible by 2."
- Then, $p \leftrightarrow q$ represents the proposition "A number is even if and only if it is divisible by 2."

- $p \leftrightarrow q$ is true in the following cases:
- If p is true (the number is even) and q is true (the number is divisible by 2).
- If p is false (the number is odd) and q is false (the number is not divisible by 2).
- $p \leftrightarrow q$ is false when p and q have different truth values:
- If p is true (the number is even) and q is false (the number is not divisible by 2).
- If p is false (the number is odd) and q is true (the number is divisible by 2).

2.2.3 Truth Tables

Truth tables are a systematic way to represent and analyze the truth values of propositions and compound propositions. The table lists all possible combinations of truth values for the constituent propositions and shows the resulting truth value of the compound proposition for each of those combinations.

2.2.3.1 Construction of Truth Tables

To construct a truth table, follow these steps:

1) Identify the constituent propositions involved in the compound proposition.
2) Determine the number of rows needed in the truth table (2^n, where n is the number of constituent propositions).
3) List all possible combinations of truth values for the constituent propositions; a systematic method for doing so will be shown by example.
4) Evaluate the truth value of the compound proposition for each combination using the definitions of the logical connectives.

Example 2.6 Construct a truth table for the compound proposition $(p \wedge q) \rightarrow r$.

1) The constituent propositions are p, q, and r.
2) There are three propositions, so the truth table will have 2^3 = 8 rows.
3) List all possible combinations of truth values for p, q, and r.
4) Evaluate the truth value of $(p \wedge q) \rightarrow r$ for each combination.

To ensure we cover all possible combinations of truth values for propositions p, q, and r, we adopt the following system:

- Use true (T) for the first half of the first column, false (F) for the second half of the first column.
- In the second column, repeat a pattern of 2^{n-1} T's, followed by 2^{n-1} F's, 2^{n-1} T's, followed by 2^{n-1} F's.
- Iterate this process until all constituent proposition columns are complete.

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

Before proceeding, review the rules that determine the truth value for proposition combinations (following the order of operations, using parentheses to determine nested expressions). The thought process runs along these lines: $(p \wedge q)$ is only true in the rows, where both p and q are true; $(p \wedge q) \rightarrow r$ is true EXCEPT those rows where $(p \wedge q)$ is true, but r is false.

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

The rationale for completing truth tables is highly important. It is, in a sense, an emulation of machine logic, and a methodology about which we must have a very strong understanding.

Example 2.7 Construct a truth table for the compound proposition $(\neg p \vee r) \wedge q \leftrightarrow (r \vee q)$

Solution: There are three constituent propositions, so the truth table will have $2^3 = 8$ rows.

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

2.2.3.2 Categorization of Truth Table Results

Truth tables can be used to analyze the logical properties of compound propositions and to identify the placement of those results into broad categories. By examining the truth values in the final column of the truth table, we can determine whether a compound proposition is:

- Always true (a tautology).
- Always false (a contradiction).
- Sometimes true and sometimes false (a contingency).

2.2.3.3 Tautology, Contradiction, and Contingency

A tautology is a compound proposition that is always true, regardless of the truth values of its constituent propositions. In a truth table, a tautology has only "T" (true) in its final column.

Example 2.8 Demonstrate the compound proposition $p \vee \neg p$ (this result is also known as the law of excluded middle) is a tautology

P	¬p	p ∨ ¬p
T	F	T
F	T	T

Since the final column contains only "T," $p \vee \neg p$ is a tautology.

As an illustration, consider the proposition p: "a whole number is even." The law of the excluded middle tells us (the obviously true fact that) a whole number is even or a whole number is not even. That is, we have a situation in which something either is in a particular category or it is not.

A contradiction is a compound proposition that is always false, regardless of the truth values of its constituent propositions. In a truth table, a contradiction has only "F" (false) in its final column.

Example 2.9 Demonstrate the compound proposition $p \wedge \neg p$ is a contradiction.

P	¬p	p ∧ ¬p
T	F	F
F	T	F

Since the final column contains only "F," $p \wedge \neg p$ is a contradiction. Appealing once more to our illustration, where the proposition p: "a whole number is even," we can see that it is impossible for a whole number to be both even and yet not even. In this situation, we can see a contradiction for what it is: an absurdity.

A contingency is a compound proposition that is neither a tautology nor a contradiction. Its truth value depends on the truth values of its constituent propositions. In a truth table, a contingency has both "T" and "F" in its final column.

Example 2.10 Demonstrate the compound proposition $p \rightarrow q$ is a contingency.

p	Q	p → q
T	T	T
T	F	F
F	T	T
F	F	T

Since the final column contains both "T" and "F," $p \to q$ is a contingency.

Imagine p: "I go to school every day," q: "My attendance record is perfect." The only situation in which this is false is this: If I do go to school every day, then my attendance record is not perfect. We can see this makes no sense.

The scenario just described raises the reasonable curiosity regarding the truth of an implication when the antecedent (that statement preceding "then" in the implication) is false, but the consequent is true. We say such a case is "vacuously true." A vacuously true statement is a logical assertion that is deemed true because the condition under which it would be false never occurs.

Example 2.11 Determine whether the compound proposition $(p \wedge q) \to r$ is a tautology, a contradiction, or a contingency.

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

By examining the final column, we can see that the compound proposition is sometimes true and sometimes false, depending on the truth values of p, q, and r. Therefore, $(p \wedge q) \to r$ is a contingency.

Example 2.12 Determine whether the compound proposition $(\neg p \vee r) \wedge q \leftrightarrow (r \vee q)$ is a tautology, a contradiction, or a contingency.

Solution: Recalling the truth table constructed earlier,

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

we see the compound proposition is a contingency.

2.3 Logical Equivalence

Logical equivalence is a fundamental concept in propositional logic that allows us to identify when two compound propositions have the same logical meaning, that is, they are true under the same conditions. Understanding logical equivalence is essential for simplifying complex logical expressions and proving the validity of arguments.

2.3.1 Definition of Logical Equivalence

Two compound propositions, P and Q, are said to be logically equivalent if they have the same truth value for every possible combination of truth values of their constituent propositions. In other words, P and Q are logically equivalent if their truth tables are identical in the final column.

We denote logical equivalence using the symbol "≡." If P and Q are logically equivalent, we write $P \equiv Q$.

Example 2.13 Demonstrate the compound propositions $P : p \wedge q$ and $Q : q \wedge p$ are logically equivalent.

p	q	P : p ∧ q	Q : q ∧ p
T	T	T	T
T	F	F	F
F	T	F	F
F	F	F	F

Since the truth tables for P and Q are identical in the final column, we can conclude that $P \equiv Q$. This logical equivalence is known as the commutative law for conjunction.

Example 2.14 Demonstrate the compound propositions $P : p \rightarrow q$ and $Q: \neg p \vee q$ are logically equivalent.

p	q	P : p → q	Q: ¬p ∨ q
T	T	T	T
T	F	F	F
F	T	T	T
F	F	T	T

Since the truth tables for P and Q are identical in the final column, we can conclude that $P \equiv Q$. This equivalence demonstrates that an implication can be expressed as a disjunction of the negation of the antecedent and the consequent.

Example 2.15 Demonstrate the compound propositions P: ¬(p ∨ q) and Q: ¬p ∧ ¬q are logically equivalent.

p	Q	¬p	¬q	P: ¬(p ∨ q)	Q: ¬p ∧ ¬q
T	T	F	F	F	F
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

Since the truth tables for P and Q are identical in the final column, we can conclude that P ≡ Q.

The equivalence in this example is a particularly important one and is known as De Morgan's law for negation of disjunction.

2.3.2 Laws of Logical Equivalence

Logical equivalence laws allow us to simplify and transform logical expressions while maintaining their logical meaning. These laws are based on the properties of logical connectives and can be used to prove the equivalence of two compound propositions. In practical terms, we can use these laws to establish whether two "versions" of a statement do, in fact, make the same claim.

2.3.2.1 Commutative Laws of Conjunction and Disjunction

The commutative laws state that the order of the propositions in a conjunction or disjunction does not affect the truth value of the compound proposition. We recall that, in algebra, not all operations possess the property of commutativity, and this is true in logic, as well.

Commutative law for conjunction: $p \wedge q \equiv q \wedge p$.

Commutative law for disjunction: $p \vee q \equiv q \vee p$.

Example 2.16 Let p be the proposition "I like apples," and q be the proposition "I like bananas."

$p \wedge q$: "I like apples and I like bananas."

$q \wedge p$: "I like bananas and I like apples."

Both compound propositions have the same meaning and truth value (we state this as a truism; construction of the truth table for each compound proposition calls merely for interchanging the order of the first two columns of the table), demonstrating the commutative law for conjunction.

2.3.2.2 Associative Laws

The associative laws state that the grouping of propositions in a conjunction or disjunction does not affect the truth value of the compound proposition.

Associative law for conjunction: $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$.

Associative law for disjunction: $(p \vee q) \vee r \equiv p \vee (q \vee r)$.

That these laws hold can be demonstrated through a truth table. We will show the result for conjunction while leaving that for disjunction as an exercise.

p	Q	r	p ∧ q	q ∧ r	(p ∧ q) ∧ r	p ∧ (q ∧ r)
T	T	T	T	T	T	T
T	T	F	T	F	F	F
T	F	T	F	F	F	F
T	F	F	F	F	F	F
F	T	T	F	T	F	F
F	T	F	F	F	F	F
F	F	T	F	F	F	F
F	F	F	F	F	F	F

Since the final two columns of the truth table are identical, we have established $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$.

Example 2.17 Let p be the proposition "I have a pen," q be the proposition "I have a pencil," and r be the proposition "I have a marker."

$(p \wedge q) \wedge r$: "(I have a pen and I have a pencil) and I have a marker."

$p \wedge (q \wedge r)$: "I have a pen and (I have a pencil and I have a marker)."

Both compound propositions have the same meaning and truth value, demonstrating the associative law for conjunction. Observe this agrees with an intuitive sense the two statements make precisely the same claim.

2.3.2.3 Distributive Laws

The distributive laws describe how conjunction and disjunction interact with each other.

Distributive law of conjunction over disjunction: $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$.

Distributive law of disjunction over conjunction: $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$.

Again, we prove the case for the distributive law of conjunction over disjunction and leave the remaining proof to the exercises.

p	q	r	p ∧ q	p ∧ r	q ∨ r	p ∧ (q ∨ r)	(p ∧ q) ∨ (p ∧ r)
T	T	T	T	T	T	T	T
T	T	F	T	F	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	F	F	F	F
F	T	T	F	F	T	F	F
F	T	F	F	F	F	F	F
F	F	T	F	F	F	F	F
F	F	F	F	F	F	F	F

Since the final two columns of the truth table are identical, we have established $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$.

Example 2.18 Let p be the proposition "I will buy a shirt," q be the proposition "The shirt is blue," and r be the proposition "The shirt is green."

$p \wedge (q \vee r)$: "I will buy a shirt and (the shirt is blue or the shirt is green)."

$(p \wedge q) \vee (p \wedge r)$: "(I will buy a shirt and the shirt is blue) or (I will buy a shirt and the shirt is green)." Both compound propositions have the same meaning and truth value, demonstrating the distributive law of conjunction over disjunction.

2.3.2.4 De Morgan's Laws

De Morgan's laws describe how negation interacts with conjunction and disjunction.

- De Morgan's law for conjunction: $\neg(p \wedge q) \equiv \neg p \vee \neg q$.
- De Morgan's law for disjunction: $\neg(p \vee q) \equiv \neg p \wedge \neg q$.

Unlike our practice of proving one of the two laws, we leave verification by truth tables to the exercises. We will continue this practice in subsequent sections of this chapter.

Example 2.19 Let p be the proposition "The car is red," and q be the proposition "The car is a sedan."

$\neg(p \wedge q)$: "It is not the case that (the car is red and the car is a sedan)."

$\neg p \vee \neg q$: "The car is not red or the car is not a sedan."

Both compound propositions have the same meaning and truth value, demonstrating De Morgan's law for conjunction.

It may be illustrative to consider a specific example pertaining to the situation just described. Imagine we have a red convertible. This satisfies the intent behind $\neg(p \wedge q)$; it is not the case that we have a vehicle that is both red and a sedan. On the other hand, the interpretation of $\neg p \vee \neg q$ would be that we have a vehicle that is not red, or is not a sedan. While $\neg p$ is false, $\neg q$ is true. A disjunction is true provided one or both of the statements involved are true.

2.3.2.5 Absorption Laws

The absorption laws describe how a proposition can be absorbed by a conjunction or disjunction with itself.

- Absorption law for conjunction: $p \wedge (p \vee q) \equiv p$.
- Absorption law for disjunction: $p \vee (p \wedge q) \equiv p$.

Example 2.20 Let p be the proposition "I will go to the park," and q be the proposition "I will go to the beach."

$p \wedge (p \vee q)$: "I will go to the park and (I will go to the park or I will go to the beach)."

The final column of the truth table for the compound proposition simplifies to the column of the truth table for p. Thus, the overall compound statement is equivalent to the simple statement "I will go to the park," demonstrating the absorption law for conjunction.

2.3.2.6 Idempotent Laws

The idempotent laws state that a proposition in conjunction or disjunction with itself is logically equivalent to the original proposition.

- Idempotent law for conjunction: $p \wedge p \equiv p$.
- Idempotent law for disjunction: $p \vee p \equiv p$.

Example 2.21 Let p be the proposition "I am studying."

$p \wedge p$: "I am studying and I am studying." The compound proposition simplifies to just "I am studying," demonstrating the idempotent law for conjunction.

2.3.2.7 Double Negation Law

The double negation law states that the negation of the negation of a proposition is logically equivalent to the original proposition.

- Double negation law: $\neg(\neg p) \equiv p$.

Example 2.22 Let p be the proposition "The sky is blue."

$\neg(\neg p)$: "It is not the case that it is not the case that the sky is blue." The compound proposition simplifies to "The sky is blue," demonstrating the double negation law.

2.3.2.8 Simplification Law

If the compound statement $p \wedge q$ is true, then both p and q are individually true.

As an illustration, imagine I have a dog named Spot, and I claim Spot is 7-year old, and Spot is a cocker spaniel. Then, by simplification, it is true to say Spot is 7-year old, and also to say Spot is a cocker spaniel.

2.4 Logical Inference and Proofs

Logical inference is the process of deriving new propositions from existing ones using logical rules and methods. In this section, we will explore the concepts of validity and soundness, as well as various inference rules and their applications in constructing proofs.

We adopt the convention of using the symbol $\therefore$ to represent the word "therefore," which precedes the conclusion.

2.4.1 Validity and Soundness

An argument is a sequence of propositions, called premises, followed by a concluding proposition, known as the conclusion. The validity and soundness of an argument are two essential properties that determine its strength and reliability.

An argument is valid if the truth of its premises necessarily leads to the truth of its conclusion. In other words, if the premises are true, the conclusion must be true.

An argument is sound if it is valid, **and** all of its premises are actually true.

Example 2.23 Consider the following argument:

All cats are mammals. (Premise 1)

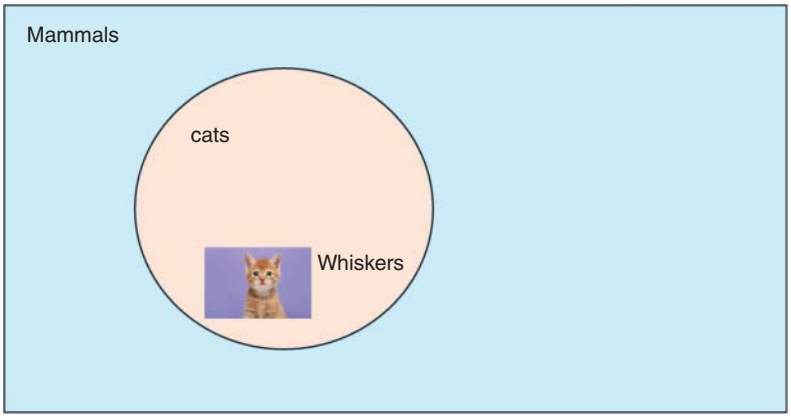
Whiskers is a cat. (Premise 2)

$\therefore$ Whiskers is a mammal. (Conclusion)

This argument is valid because if both premises are true, the conclusion must be true.

Additionally, if we know that both premises are factually true, then the argument is also sound.

For a relatively simple argument such as this, a Venn Diagram is useful to illustrate the situation:



Source: Elles Rijsdijk / Adobe Stock Photos.

Clearly, since the set of all cats lies within the universal set of all mammals, and Whiskers is within the set of all cats, Whiskers is a mammal.

2.4.2 Modus Ponens and Modus Tollens

Modus ponens and modus tollens are two fundamental inference rules in propositional logic.

Modus ponens, also known as the law of detachment, states that if p is true and $p \rightarrow q$ is true, then q must be true.

Symbolically: $p, p \rightarrow q \therefore q$.

Example 2.24 If it is raining, the ground is wet. $(p \rightarrow q)$

It is raining. (p)

$\therefore$ The ground is wet. (q)

Consider the illustration in a truth table:

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

Observe q is true in the event both p and $p \rightarrow q$ are true.

Modus tollens, also known as the law of contrapositive, states that if $\neg q$ is true and $p \rightarrow q$ is true, then $\neg p$ must be true.

Symbolically: $\neg q, p \rightarrow q \therefore \neg p$.

Example 2.25 If the light is on, the switch is closed. $(p \rightarrow q)$

The switch is not closed. $(\neg q)$

$\therefore$ The light is not on. $(\neg p)$

Consider the illustration in a truth table:

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

Observe that when both $\neg q$ and $p \to q$ are true, $\neg p$ is true.

2.4.3 Hypothetical Syllogism

Hypothetical syllogism is an inference rule that combines two conditional statements to form a new conditional statement. This is sometimes referred to as a "chain of implications."

Symbolically: $p \to q$, $q \to r \therefore p \to r$.

Example 2.26 If I study hard, I will pass the exam. $(p \to q)$
If I pass the exam, I will graduate. $(q \to r)$
$\therefore$ If I study hard, I will graduate. $(p \to r)$

2.4.4 Disjunctive Syllogism

Disjunctive syllogism is an inference rule that allows us to conclude one disjunct when the other disjunct is negated.

Symbolically: $p \vee q$, $\neg p \therefore q$ (or $p \vee q$, $\neg q \therefore p$).

Example 2.27 The car is either red or blue. $(p \vee q)$
The car is not red. $(\neg p)$
$\therefore$ The car is blue. (q)

This form of argument is sometimes referred to as "ruling out the impossible."

2.4.5 Constructing Proofs Using Inference Rules

Proofs are logical arguments that demonstrate the validity of a conclusion based on a set of premises. By applying inference rules, we can construct proofs to show that a conclusion follows from given premises.

Example 2.28 Prove that the conclusion q follows from the premises $p \to (q \wedge r)$ and p.

1) $p \to (q \wedge r)$ (Premise).
2) p (Premise).
3) $q \wedge r$ (Modus ponens, 1, 2).
4) q (Simplification, 3).

In this proof, we first use modus ponens to derive $q \wedge r$ from the given premises. Then, we use the simplification rule (which states that if a conjunction is true, each of its conjuncts is true) to conclude q.

2.5 Logic Gates

Logic gates are fundamental building blocks in digital electronics and computer architecture. They are used to perform logical operations on binary inputs and thereby produce binary outputs. Understanding logic gates is essential for designing and analyzing digital circuits and systems.

2.5.1 Introduction to Logic Gates

A logic gate is an electronic circuit that implements a Boolean function (a function having n variables or entries, so it has 2^n possible combinations of the variables), taking one or more binary inputs and producing a single binary output. The inputs and outputs of logic gates are represented by two voltage levels: a low voltage (usually 0 V) representing a logical "0" or "false," and a high voltage (usually 3.3 V or 5 V) representing a logical "1" or "true."

There are three basic logic gates: AND, OR, and NOT. These gates can be combined to create more complex gates and circuits.

AND Gate:

An AND gate outputs a logical "1" only if all its inputs are logical "1"; otherwise, it outputs a logical "0."

The Boolean expression for an AND gate with inputs A and B is: $Y = A \wedge B$

Truth table for a two-input AND gate:

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

Example 2.29 A simple alarm system uses an AND gate to trigger the alarm. The inputs are "motion detected" and "system armed." The alarm will sound only if both inputs are true.

OR Gate:

An OR gate outputs a logical "1" if at least one of its inputs is logical "1"; otherwise, it outputs a logical "0."

The Boolean expression for an OR gate with inputs A and B is: $Y = A \vee B$.

Truth table for a two-input OR gate:

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

Example 2.30 A car's brake light system uses an OR gate. The inputs are "brake pedal pressed" and "emergency brake engaged." The brake lights will turn on if either input is true.

NOT Gate (Inverter):

A NOT gate, also known as an inverter, outputs the logical complement of its input. If the input is logical "1," the output is logical "0," and vice versa.

The Boolean expression for a NOT gate with input A is: $Y = \neg A$.

Truth table for a NOT gate:

A	Y
0	1
1	0

Example 2.31 A simple cooling system uses a NOT gate to control a fan. The input is "temperature below threshold." When the temperature is above the threshold (input is false), the fan turns on (output is true).

2.5.2 Types of Logic Gates

The basic logic gates can be combined to form more complex gates, such as:

NAND (NOT–AND) Gate: $Y = \neg(A \wedge B)$.
NOR (NOT–OR) Gate: $Y = \neg(A \vee B)$.
XOR (Exclusive-OR) Gate: $Y = (A \wedge \neg B) \vee (\neg A \wedge B)$.
XNOR (Exclusive-NOR) Gate: $Y = (A \wedge B) \vee (\neg A \wedge \neg B)$.

2.5.2.1 NAND Gate

A NAND (NOT-AND) gate is a combination of an AND gate followed by a NOT gate. It outputs a "0" only if all its inputs are "1." In all other cases, the output is "1." The symbol for a NAND gate is an AND gate with a bar over the output or the word "NAND."

Boolean expression: $Y = \neg(A \cdot B)$ or $Y = (A \wedge B)'$.

Truth table for a two-input NAND gate:

A	B	Y
0	0	1
0	1	1
1	0	1
1	1	0

Example 2.32 A factory uses a NAND gate to control an automated assembly line. The inputs are "safety sensor 1 activated" (A) and "safety sensor 2 activated" (B). The assembly line (Y) will stop only if both safety sensors are activated.

Safety Sensor 1 (A)	Safety Sensor 2 (B)	Assembly Line Running (Y)
0	0	1
0	1	1
1	0	1
1	1	0

The NAND gate is considered a universal gate because any other logic gate (AND, OR, and NOT) can be constructed using only NAND gates.

2.5.2.2 NOR Gate

A NOR (NOT–OR) gate is a combination of an OR gate followed by a NOT gate. It outputs a "1" only if all its inputs are "0." In all other cases, the output is "0." The symbol for a NOR gate is an OR gate with a bar over the output or the word "NOR."

Boolean expression: $Y = \neg(A + B)$ or $Y = (A \vee B)'$.

Truth table for a two-input NOR gate:

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	0

Example 2.33 A car's automatic transmission uses a NOR gate to determine when to shift into high gear. The inputs are "engine speed < 2000 rpm" (A) and "throttle position < 50%" (B). The transmission will shift into high gear (Y) only if both conditions are false.

Engine Speed <2000 rpm (A)	Throttle Position <50% (B)	Shift into High Gear (Y)
0	0	1
0	1	0
1	0	0
1	1	0

Like the NAND gate, the NOR gate is also a universal gate, meaning that any other logic gate can be constructed using only NOR gates.

2.5.2.3 XOR Gate

An XOR (Exclusive-OR) gate outputs a "1" when the number of "1" inputs is odd. In other words, the output is "1" if exactly one input is "1." If both inputs are "0" or both inputs are "1," the output is "0." The symbol for an XOR gate is a plus sign inside a circle or the word "XOR."

Boolean expression: $Y = A \oplus B$ or $Y = (A \wedge \neg B) \vee (\neg A \wedge B)$.

Truth table for a two-input XOR gate:

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Example 2.34 A security system uses an XOR gate to compare the entered password with the stored password. The inputs are the corresponding bits of the entered password (A) and the stored password (B). The output (Y) is "1" if the bits are different, indicating an incorrect password.

Entered Password Bit (A)	Stored Password Bit (B)	Output (Y)
0	0	0
0	1	1
1	0	1
1	1	0

XOR gates are commonly used in error detection and correction circuits, such as parity checkers and arithmetic circuits.

2.5.2.4 XNOR Gate

An XNOR (Exclusive-NOR) gate, also known as an "equivalence" gate, outputs a "1" when the number of "1" inputs is even. In other words, the output is "1" if both inputs are the same (either both "0" or both "1"). The output is "0" if the inputs are different. The symbol for an XNOR gate is an XOR gate with a bar over the output or the word "XNOR." That is, we would see either $(\overline{A \oplus B})$ or A XNOR B. An alternative notation is $A \odot B$. Boolean expression: $Y = A \equiv B$ or $Y = \neg(A \oplus B)$ or $Y = (A \wedge B) \vee (\neg A \wedge \neg B)$.

Truth table for a two-input XNOR gate:

A	B	Y
0	0	1
0	1	0
1	0	0
1	1	1

Example 2.35 A circuit uses an XNOR gate to compare two binary numbers, A and B. The output (Y) is "1" if the numbers are equal and "0" if they are different.

Binary Number A	Binary Number B	Equal (Y)
00	00	1
01	00	0
10	10	1
11	01	0

XNOR gates are often used in circuits that check for equality, such as comparators and coincidence detectors.

2.5.3 Combinational Logic Circuits

Combinational logic circuits are digital circuits in which the output depends solely on the current input values. These circuits are, naturally, constructed using combinations of various logic gates. The output of a combinational logic circuit is determined by the logical operations performed on the inputs, without any memory of previous input states.

Examples of combinational logic circuits include:

Half Adder: A half adder is a combinational logic circuit that performs the addition of two single-bit binary numbers (A and B) and produces a sum (S) and a carry (C) output.

Boolean expressions:

- $S = A \oplus B$ (sum).
- $C = A \wedge B$ (carry).

> Note: It may seem troublesome we appear to be reusing symbols for XNOR and disjunction at this point. Use of truth tables abates this concern, however: the truth tables for XNOR and sum, and the truth tables for disjunction and carry turn out to be identical.

Truth table for a half adder:

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

The half adder can be implemented using an XOR gate for the sum and an AND gate for the carry.

Full Adder:

A full adder is a combinational logic circuit that performs the addition of three single-bit binary numbers: two input bits (A and B) and a carry-in bit (Cin). It produces a sum (S) and a carry-out (Cout) output.

Boolean expressions:

- $S = A \oplus B \oplus Cin$ (sum).
- $Cout = (A \wedge B) \vee (Cin \wedge (A \oplus B))$ (carry-out).

Truth table for a full adder:

A	B	Cin	S	Cout
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

A full adder can be implemented using two half adders and an OR gate. The sum output is obtained by connecting the sum outputs of the half adders to an XOR gate, while the carry-out is the output of the OR gate, which takes the carry outputs of the half adders as inputs.

An illustration is useful here as motivation for the terminology. We know that, when adding base 10 numbers by hand, we use a sum and carry procedure. That is, we add down columns of the base 10 numbers and have to "carry a 1" to the next column if the sum of column entries exceeds 9.

A similar process occurs in the addition of binary numbers, but the process is more "extreme."

There are four binary sums: $0+0=0$, $0+1=1$, $1+0=0$, and $1+1=10$. Observe the addition of the two single-digit 1's produces a two-digit binary number. That is, the addition of a single bit produces a two-bit output.

When you want to sum inputs with more than 1-bit numbers, for example, adding two 4-bit binary numbers, you will need to combine several adders and provide a carry input between them to express that the previous less significant operation produced a carry.

2-to-1 Multiplexer (MUX):

A 2-to-1 multiplexer is a combinational logic circuit that selects one of two input signals (A or B) based on a select input (S) and outputs the selected signal (Y). This is a fundamental circuit, which is used to choose one signal from two inputs and transmit it to the output. The 2-to-1 MUX has two input lines, one output line, and a single selection line. It has various applications in digital systems, such as in a microprocessor, where it is used to select between two different data sources or between two different instructions.

Boolean expression:

- $Y = (\neg S \wedge A) \vee (S \wedge B)$.

Truth table for a 2-to-1 MUX:

S	A	B	Y
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

A 2-to-1 MUX can be implemented using AND, OR, and NOT gates, as shown in the Boolean expression.

2.5.4 Sequential Logic Circuits

Sequential logic circuits are digital circuits in which the output depends not only on the current input values but also on the previous state of the circuit. Unlike combinational logic circuits, sequential logic circuits have memory elements that store the state information. The output of a sequential logic circuit is determined by the current inputs and the stored state. The most common memory elements used in sequential logic circuits are flip-flops and latches.

SR (Set–Reset) Latch:

An SR latch is a simple sequential logic circuit that stores one bit of information. It has two inputs: Set (S) and Reset (R), and two outputs: Q and its complement Q′. The latch is set when S = 1 and

R = 0, and it is reset when S = 0 and R = 1. When both S and R are 0, the latch maintains its previous state (memory). The condition S = 1 and R = 1 is avoided, as it leads to an invalid state.

Truth table for an SR latch:

S	R	Q	Q′
0	0	Q(previous)	Q′(previous)
0	1	0	1
1	0	1	0
1	1	Invalid	Invalid

An SR latch can be implemented using two cross-coupled NOR gates or two cross-coupled NAND gates.

Clarification is needed regarding the unusual results in the truth table shown above. What do we mean, for example, by "Q(previous)" and "Q' (previous)?"

The SR latch has a stored input and thus provides a form of feedback. That is, the outcome of one "loop" of the process has an output of 1 or 0; Q(previous) is simply the output from the previous loop of the procedure.

Why is the input pairing S = R = 1 "invalid?"

Keep in mind the SR latch is a set-reset switch. If both inputs are 1, then we would be telling the switch simultaneously to set itself and to reset itself, a state which is highly undesirable. Doing so would produce a situation where both Q and Q′ would be 0. This could sort itself out if one of the inputs were to fall to zero before the other. If both inputs fell to zero simultaneously, we would encounter a "race condition," where each gate would be competing to see which would produce its output first. It would be impossible to reliably determine which output would "win" this race, and so the next state of the latch could not be guaranteed.

D (Data) Flip-Flop:

A D flip-flop is a sequential logic circuit that stores one bit of information and is triggered by a clock signal. It has one data input (D), one clock input (CLK), and two outputs: Q and its complement Q′. The output Q takes on the value of the input D at the rising edge of the clock signal.

Truth table for a D flip-flop:

CLK	D	Q
Rising edge	0	0
Rising edge	1	1

A D flip-flop can be implemented using an SR latch and additional logic gates to ensure that the S and R inputs are never both 1.

Shift Register:

A shift register is a sequential logic circuit that consists of a series of flip-flops connected in a chain. It is used to store and shift binary data. Each flip-flop in the shift register stores one bit of data, and the data is shifted from one flip-flop to the next on each clock cycle. Shift registers can be used for serial-to-parallel and parallel-to-serial data conversion, data delay, and temporary data storage.

Example 2.36 A 4-bit shift register with a serial input (SI), a serial output (SO), and a clock input (CLK).

CLK	SI	Q3	Q2	Q1	Q0	SO
Rising edge	1	0	0	0	1	0
Rising edge	0	0	0	1	0	0
Rising edge	1	0	1	0	1	1
Rising edge	1	1	0	1	1	0

A shift register can be implemented using a series of D flip-flops, where the output (Q) of each flip-flop is connected to the input (D) of the next flip-flop in the chain.

2.6 Predicate Logic

Predicate logic, also known as first-order logic, is an extension of propositional logic that allows for more expressive statements. While propositional logic deals with simple declarative statements, predicate logic introduces the concepts of predicates, variables, and quantifiers, enabling us to represent and reason about more complex mathematical and real-world situations.

2.6.1 Predicates and Quantifiers

A predicate is a statement that contains variables and expresses a property or relationship between those variables. Unlike propositions, which are either true or false, predicates become true or false only when specific values are assigned to their variables. Predicates are usually denoted by uppercase letters, such as P(x), Q(x, y), or R(x, y, z).

Example 2.37 Let P(x) be the predicate "x is even," where x is a variable representing an integer.
P(4) is true because 4 is an even number.
P(7) is false because 7 is not an even number.
Quantifiers are logical symbols that specify the quantity or extent of the variables in a predicate.

There are two main types of quantifiers:

- Universal quantifier (∀): It means "for all" or "for every."
- Existential quantifier (∃): It means "there exists" or "for some."

2.6.2 Universal Quantifier

The universal quantifier (∀) is used to state that a predicate is true for all values of the variable within a given domain. The domain is the set of all possible values that the variable can take. The universal quantifier is read as "for all" or "for every."

Example 2.38 Let R(x) be the predicate "x^2 ≥ 0," where x is a variable representing a real number.

∀ x R(x) means "for all real numbers x, x^2 ≥ 0," which is true because the square of any real number is always non-negative.

Example 2.39 Let S(x) be the predicate "x is a prime number," where x is a variable representing an integer.

∀x S(x) means "for all integers x, x is a prime number," which is false because not every integer is a prime number (e.g., 1, 4, 6, 8, etc., are not prime numbers).

Establishing a universally qualified statement is false can be accomplished by production of at least one situation in which the statement fails to hold. We refer to such a situation as a counterexample.

Example 2.40 Let T(x, y) be the predicate x–y = y–x, where x and y are variables representing real numbers. Show the statement

∀x ∀y T(x, y)

is false.

Solution: Considering the pair of real numbers 1 and 2, we see 1 – 2 = −1, while 2 – 1 = 1. This serves as a counterexample, and so the claim is false.In predicate logic, the universal quantifier is often used to express mathematical statements, rules, or properties that hold for all elements in a given set or domain.

2.6.3 Existential Quantifier

The existential quantifier (∃) is used to state that a predicate is true for at least one value of the variable within a given domain. The existential quantifier is read as "there exists" or "for some."

Example 2.41 Let P(x) be the predicate "x is a positive integer less than 10," where x is a variable representing an integer.

∃x P(x) means "there exists an integer x such that x is a positive integer less than 10," which is true (e.g., 1, 2, 3, ..., 9 satisfy the predicate).

Example 2.42 Let Q(x, y) be the predicate "x^2 + y^2 = 25," where x and y are variables representing real numbers.

∃x ∃y Q(x, y) means "there exist real numbers x and y such that x^2 + y^2 = 25," which is true (e.g., x = 3 and y = 4 satisfy the predicate).

Example 2.43 Let R(x) be the predicate "x is a solution to the equation x^2 + 1 = 0," where x is a variable representing a real number.

∃x R(x) means "there exists a real number x such that x is a solution to the equation x^2 + 1 = 0," which is false because the equation has no real solutions (the square of a real number is always non-negative, so x^2 + 1 cannot equal 0).

2.6.4 Negation of Quantified Statements

Negating a quantified statement involves changing the quantifier and the predicate. The negation of a universally quantified statement is an existentially quantified statement with the negated predicate, and vice versa.

- Negation of $\forall x\, P(x)$ is $\exists x\, \neg P(x)$.
- Negation of $\exists x\, P(x)$ is $\forall x\, \neg P(x)$.

Example 2.44 Consider the statement "All cats are black," which can be represented using predicate logic as $\forall x\, (C(x) \rightarrow B(x))$, where $C(x)$ means "x is a cat" and $B(x)$ means "x is black."

The negation of this statement is "There exists a cat that is not black," which can be represented as $\exists x\, (C(x) \wedge \neg B(x))$.

Example 2.45 Let $S(x)$ be the predicate "x is a student," and let $T(x)$ be the predicate "x is taking a math course," where x is a variable representing a person.

The statement "Every student is taking a math course" can be represented as $\forall x\, (S(x) \rightarrow T(x))$.

The negation of this statement is "There exists a student who is not taking a math course," which can be represented as $\exists x\, (S(x) \wedge \neg T(x))$.

Example 2.46 Let $P(x)$ be the predicate "x is a prime number," where x is a variable representing an integer.

The statement "There exists a prime number that is even" can be represented as $\exists x\, (P(x) \wedge E(x))$, where $E(x)$ means "x is even."

The negation of this statement is "For all prime numbers x, x is not even," which can be represented as $\forall x\, (P(x) \rightarrow \neg E(x))$.

2.6.5 Translating English Sentences into Predicate Logic

Translating English sentences into predicate logic is an essential skill for representing and analyzing statements using formal logic. By breaking down the sentences into their constituent parts and identifying the predicates, variables, and quantifiers, we can create precise logical representations of the original statements.

Example 2.47 "All dogs are mammals."

Let $D(x)$ be the predicate "x is a dog," and let $M(x)$ be the predicate "x is a mammal."

The given statement can be translated into predicate logic as: $\forall x\, (D(x) \rightarrow M(x))$, which reads as "for all x, if x is a dog, then x is a mammal."

Example 2.48 "Some birds can fly."

Let $B(x)$ be the predicate "x is a bird," and let $F(x)$ be the predicate "x can fly."

The given statement can be translated into predicate logic as: $\exists x\, (B(x) \wedge F(x))$, which reads as "there exists an x such that x is a bird and x can fly."

Example 2.49 "No reptile is warm-blooded."

Let $R(x)$ be the predicate "x is a reptile," and let $W(x)$ be the predicate "x is warm-blooded."

The given statement can be translated into predicate logic as: $\forall x\, (R(x) \rightarrow \neg W(x))$, which reads as "for all x, if x is a reptile, then x is not warm-blooded."

Example 2.50 "Every integer is either even or odd."

Let $I(x)$ be the predicate "x is an integer," $E(x)$ be the predicate "x is even," and $O(x)$ be the predicate "x is odd."

The given statement can be translated into predicate logic as: $\forall x\ (\mathrm{I}(x) \rightarrow (\mathrm{E}(x) \vee \mathrm{O}(x)))$, which reads as "for all x, if x is an integer, then x is even or x is odd."

Example 2.51 "There exists a prime number that is greater than 100."

Let $\mathrm{P}(x)$ be the predicate "x is a prime number," and let $\mathrm{G}(x)$ be the predicate "x is greater than 100."

The given statement can be translated into predicate logic as: $\exists x\ (\mathrm{P}(x) \wedge \mathrm{G}(x))$, which reads as "there exists an x such that x is a prime number and x is greater than 100."

Example 2.52 "For every real number x, there exists a real number y such that $x + y = 0$."

Let $\mathrm{R}(x)$ be the predicate "x is a real number."

The given statement can be translated into predicate logic as: $\forall x\ (\mathrm{R}(x) \rightarrow \exists \mathrm{y}\ (\mathrm{R[y]} \wedge x + y = 0))$, which reads as "for all x, if x is a real number, then there exists a y such that y is a real number and $x + y = 0$."

When translating English sentences into predicate logic, it is important to:

- Identify the predicates and clearly define what they represent.
- Determine the variables and their domains (e.g., integers, real numbers, or a specific set).
- Identify the quantifiers (universal or existential) and their scope.
- Break down the sentence into its constituent parts and translate each part using the appropriate logical connectives (e.g., $\wedge$, $\vee$, $\rightarrow$, $\neg$).

2.7 Applications of Logic

Logic has numerous applications across various fields, ranging from mathematics and computer science to philosophy and linguistics. In this section, we will explore some of the key areas where logical principles and techniques are applied to solve real-world problems and develop innovative technologies.

2.7.1 Digital Systems and Boolean Algebra

Digital systems, such as computers, smartphones, and other electronic devices, rely heavily on the principles of logic and Boolean algebra. Boolean algebra, named after the mathematician George Boole, is a branch of algebra that deals with the manipulation of logical expressions involving binary variables (i.e., variables that can only take on two values: 0 or 1, true or false).

In digital systems, Boolean algebra is used to design and optimize logic circuits, which are the building blocks of these systems. Logic circuits are composed of logic gates, such as AND, OR, NOT, NAND, NOR, XOR, and XNOR gates, which perform specific logical operations on binary inputs to produce binary outputs.

Example 2.53 Design a logic circuit that takes two binary inputs, A and B, and produces an output Y that is 1 only when both A and B are 1.

Solution: This can be achieved using an AND gate. The Boolean expression for the output Y is:

$\mathrm{Y} = \mathrm{A} \wedge \mathrm{B}$.

A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

Example 2.54 Design a logic circuit that takes three binary inputs, A, B, and C and produces an output Y that is 1 when the majority of the inputs are 1.

Solution: This can be achieved using a combination of AND and OR gates. The Boolean expression for the output Y is:

$$Y = (A \wedge B) \vee (B \wedge C) \vee (A \wedge C).$$

A	B	C	Y
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Boolean algebra is also used to simplify and optimize logic circuits. By applying various Boolean laws and theorems, such as De Morgan's laws, the distributive law, and the absorption law, complex logical expressions can be reduced to simpler forms, leading to more efficient and compact circuit designs.

Example 2.55 Simplify the Boolean expression: $(A \wedge B) \vee (\neg A \wedge B)$.

Solution: Using the distributive law and the complement law, we can simplify the expression as follows:

$$\begin{aligned}(A \wedge B) \vee (\neg A \wedge B) &= B \wedge (A \vee \neg A) \\ &= B \wedge 1 \\ &= B\end{aligned}$$

Therefore, the simplified expression is just B.

2.7.2 Computer Programming and Logical Operations

Logic plays a crucial role in computer programming, as it forms the basis for the design and implementation of algorithms, data structures, and control flow mechanisms. Logical operations, such as AND, OR, NOT, and XOR, are fundamental building blocks in programming languages and are used to make decisions, control program flow, and manipulate data.

In programming, logical operations are often represented by specific symbols or keywords, depending on the language. For example:

AND: &&, and, &
OR: ||, or, |
NOT: !, not, ~
XOR: ^, xor

These logical operations are used in combination with conditional statements (e.g., if, else if, else) and loops (e.g., while, for) to create complex program logic and algorithms.

Example 2.56 C++ program to determine if a number is divisible by both 2 and 3.

```
#include <iostream>
using namespace std;
int main() {
 int num;
 cout ≪ "Enter a number: ";
 cin ≫ num;
 if (num % 2 == 0 && num % 3 == 0) {
 cout ≪ num ≪ " is divisible by both 2 and 3." ≪ endl;
 } else {
 cout ≪ num ≪ " is not divisible by both 2 and 3." ≪ endl;
 }
 return 0;
}
```

In this example, the logical AND operation (&&) is used to check if the number is divisible by both 2 and 3. The program will only print the message "num is divisible by both 2 and 3" if both conditions are true.

Example 2.57 Python program to check if a character is a vowel or a consonant.

```
char = input("Enter a character: ")
if char in 'aeiouAEIOU':
 print(f"{char} is a vowel.")
else:
 print(f"{char} is a consonant.")
```

In this example, the logical OR operation is implicitly used when checking if the character is in the string "aeiouAEIOU." If the character is any one of the vowels, the program will print "char is a vowel."

Example 2.58 Java program to implement the XOR encryption algorithm.

```
public class XOREncryption {
 public static String encrypt(String message, String key) {
 StringBuilder encrypted = new StringBuilder();
 for (int i = 0; i < message.length(); i++) {
 char c = message.charAt(i);
 char k = key.charAt(i % key.length());
 encrypted.append((char) (c ∧ k));
 }
 return encrypted.toString();
 }
 public static void main(String[] args) {
 String message = "Hello, World!";
 String key = "Secret";
 String encryptedMessage = encrypt(message, key);
 System.out.println("Original message: " + message);
 System.out.println("Encrypted message: " + encryptedMessage);
 System.out.println("Decrypted message: " + encrypt(encryptedMessage, key));
 }
}
```

In this example, the XOR operation (^) is used to encrypt and decrypt the message. The XOR operation has the property that A ^ B ^ B = A, which makes it useful for simple encryption tasks. The same function is used for both encryption and decryption, as the XOR operation is its own inverse.

2.7.3 Artificial Intelligence and Logical Reasoning

Artificial intelligence (AI) is a field that focuses on creating intelligent machines that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, and language translation. Logical reasoning is a fundamental aspect of AI, as it enables machines to draw conclusions, make decisions, and solve problems based on available information and logical rules.

It is important to keep in mind AI simulates human intelligence. These are not one and the same.

One of the key applications of logic in AI is in the development of expert systems, which are computer programs designed to simulate the decision-making ability of a human expert in a specific domain. Expert systems use a knowledge base, which contains a set of logical rules and facts, and an inference engine, which applies the rules to the facts to draw conclusions and provide recommendations.

Example 2.59 A simple expert system for diagnosing car problems.

Knowledge base:

Rule 1: If the car will not start and the lights do not work, then the problem is a dead battery.
Rule 2: If the car will not start and the lights work, then the problem is a starter motor issue.
Rule 3: If the car starts but overheats, then the problem is a coolant issue.

Facts:

Fact 1: The car will not start.
Fact 2: The lights work.

Inference process:

The inference engine checks Rule 1 and finds that it does not apply because Fact 2 contradicts the second condition of the rule.

The inference engine checks Rule 2 and finds that it applies because both conditions (the car will not start and lights work) are satisfied by the facts.
The expert system concludes that the problem is a starter motor issue and provides this recommendation to the user.
Another important application of logic in AI is in the field of knowledge representation and reasoning (KR&R). KR&R focuses on representing knowledge in a form that can be efficiently processed by machines and using logical reasoning techniques to infer new knowledge from existing facts and rules.

Example 2.60 Using first-order logic to represent knowledge and reason about family relationships.

Knowledge base:
Parent(Alice, Bob)
Parent(Bob, Charlie)
Parent(Bob, David)
Sibling(x, y): = Parent(z, x) ∧ Parent(z, y) ∧ $x \neq y$
Query: Sibling(Charlie, David)
Reasoning process:
The reasoner identifies the rule for Sibling(x, y) and tries to find values for x, y, and z that satisfy the conditions.
The reasoner finds that Parent(Bob, Charlie) and Parent(Bob, David) are both true, and Charlie ≠ David.
The reasoner substitutes z = Bob, x = Charlie, and y = David into the rule and concludes that Sibling(Charlie, David) is true.

Logic is also used in AI for planning and decision-making tasks. Automated planning systems use logical representations of actions, states, and goals to generate sequences of actions that achieve a desired goal state. Decision-making systems use logical rules and probabilistic reasoning to make optimal choices under uncertainty.

Example 2.61 A simple planning problem in the blocks world domain.

Initial state: On(A, Table) ∧ On(B, Table) ∧ Clear(A) ∧ Clear(B)
Goal state: On(A, B) ∧ On(B, Table)
Actions:
Move(x, y, z): = On(x, y) ∧ Clear(x) ∧ Clear(z) → On(x, z) ∧ Clear(y) ∧ ¬On(x, y)
Planning process:
The planner identifies that the goal state requires A to be on B and B to be on the Table.
The planner finds that the Move action can achieve On(A, B) by moving A from the Table to B.
The planner generates the plan: Move(A, Table, B).

The application of logic in AI has led to significant advancements in various domains, such as natural language processing, computer vision, robotics, and autonomous systems.

2.7.4 Philosophical Arguments and Logical Fallacies

Logic plays a central role in philosophical arguments, as it provides a framework for analyzing and evaluating the validity and soundness of arguments. Philosophers use logical principles to construct and critique arguments and to identify and avoid logical fallacies, which are errors in reasoning that undermine the validity of an argument.

A valid argument is one in which the conclusion necessarily follows from the premises, assuming the premises are true. A sound argument is a valid argument with true premises.

Philosophical arguments aim to be both valid and sound, as this ensures that the conclusion is not only logically consistent with the premises but also true.

Example 2.62 A valid and sound argument.

Premise 1: All humans are mortal.

Premise 2: Socrates is a human.

Conclusion: Therefore, Socrates is mortal.

This argument is valid because the conclusion follows logically from the premises. If the premises are true, then the conclusion must also be true. The argument is also sound because both premises are actually true.

However, not all arguments that appear to be valid are actually sound. An argument can be valid but unsound if one or more of its premises are false.

Example 2.63 A valid but unsound argument.

Premise 1: All birds can fly.

Premise 2: Penguins are birds.

Conclusion: Therefore, penguins can fly.

This argument is valid because the conclusion follows logically from the premises. However, it is unsound because Premise 1 is false—not all birds can fly (e.g., penguins, ostriches, and kiwis).

Logical fallacies are errors in reasoning that can make arguments appear convincing, even though they are invalid or unsound.

Ad hominem: Attacking the character of the person making the argument, rather than the argument itself.

Example 2.64 "You cannot trust John's argument about climate change because he is a known liar."

Appeal to authority: Claiming that an argument is true because an authority figure believes it, even if the authority's expertise is not relevant to the argument.

Example 2.65 "The Pope believes in God, so God must exist."

False dilemma: Presenting a limited set of options as if they were the only possibilities, when in fact there are other alternatives.

Example 2.66 "Either we invest in nuclear power, or we will run out of energy. We don't want to run out of energy, so we must invest in nuclear power."

Hasty generalization: Drawing a conclusion based on insufficient or unrepresentative evidence.

Example 2.67 "I met two rude French people on my vacation. Therefore, all French people must be rude."

Slippery slope: Arguing that a small step will inevitably lead to a chain of events culminating in a significant and usually negative effect, without sufficient evidence to support the claim.

Example 2.68 "If we allow same-sex marriage, it will lead to people marrying their pets, and eventually the collapse of civilization."

Philosophers use their understanding of logic and logical fallacies to construct sound arguments and to identify and refute flawed arguments. By applying logical principles to philosophical questions, such as the nature of reality, the existence of God, or the foundations of morality, philosophers aim to arrive at well-reasoned and defensible conclusions.

2.8 Summary

Propositions: Statements that are either true or false, but not both.

Logical connectives: Operators used to combine propositions, such as:

Negation (¬): "not"

Conjunction (∧): "and"

Disjunction (∨): "or"

Implication (→): "if...then"

Biconditional (↔): "if and only if"

Truth tables: Tables used to represent the truth values of propositions and the results of logical operations.

Tautology, contradiction, and contingency: A tautology is a proposition that is always true, a contradiction is always false, and a contingency can be either true or false, depending on the truth values of its components.

Logical equivalence: Two propositions are logically equivalent if they have the same truth values for all possible combinations of their components.

Predicate logic: An extension of propositional logic that includes predicates, quantifiers, and variables.

Predicates: Functions that map objects to truth values.

Quantifiers: Symbols used to express the quantity of objects that satisfy a predicate (∀ for "for all" and ∃ for "there exists").

Logical inference: The process of deriving new propositions from existing ones using rules of inference, such as:

Modus ponens: If $P \rightarrow Q$ is true, and P is true, then Q is true.

Modus tollens: If $P \rightarrow Q$ is true, and Q is false, then P is false.

Proofs: Step-by-step demonstrations of the validity of a logical argument using rules of inference and logical equivalences.

Applications of logic: Logic is used in various fields, such as:

Digital systems and Boolean algebra

Computer programming and logical operations

Artificial intelligence and logical reasoning

Philosophical arguments and logical fallacies

3

Introduction to Number Systems

In this chapter, we will explore the important world of number systems, focusing on binary and other commonly used number systems in mathematics, cybersecurity, information systems, and computer science. We will begin by defining a number system and expressing why it is crucial in these fields.

3.1 What Is a Number System?

A number system is a way of representing numbers using a set of symbols called digits. The most familiar number system is the decimal system, which uses 10 digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) to represent numbers. However, there are many other number systems, each with its own set of digits and rules for representing and manipulating numbers.

The number of digits used in a particular number system is called its base or radix. For example, the decimal system is a base-10 system because it uses 10 digits. Other common number systems include:

- Binary (base-2): Uses digits 0 and 1.
- Octal (base-8): Uses digits 0, 1, 2, 3, 4, 5, 6, and 7.
- Hexadecimal (base-16): Uses digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and letters A, B, C, D, E, and F to represent values 10 through 15.

Example 3.1 The decimal number 42 can be represented in different number systems:

Binary: 101010
Octal: 52
Hexadecimal: 2A

At this point, it is not expected that you recognize the binary, octal, and hexadecimal representations of the decimal number 42; this will be made clear shortly. As we shall see, understanding the various representations involves careful attention to the meaning of "place names" in the various number systems, as well as the terminology we may have taken for granted in our practical experience with base-10.

Perhaps most perplexing in the introduction of alternative number systems is the omission of particular digits with which we are so familiar.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

3.1.1 Importance of Number Systems in Discrete Math and Computer Science

Number systems play a crucial role in discrete mathematics (the study of mathematical structures that are countable or otherwise distinct and separable) and computer science for several reasons

- Computer hardware: Computers use the binary number system to represent and process data because electronic circuits can easily distinguish between two states: on (1) and off (0). This makes binary a natural choice for computer systems.
- Digital systems: Number systems are essential in designing and analyzing digital circuits, such as logic gates and digital adders, which form the building blocks of modern computers and digital devices.
- Data representation: Different number systems are used to represent various types of data efficiently. For example, hexadecimal is often used to represent colors in web design (e.g., #FF0000 represents red) and memory addresses in programming. Algorithms and data structures: Understanding number systems is crucial for developing efficient algorithms and data structures, especially those involving bit manipulation or low-level system programming.

Example 3.2 In computer networking, IPv4 addresses are typically represented using dotted-decimal notation (e.g., 192.168.0.1). However, these addresses are actually 32-bit binary numbers. Understanding binary and hexadecimal number systems is essential for working with IP addresses and other low-level networking concepts.

3.1.2 Place Value and Positional Notation in a Number System

Before commencing our exploration of alternative number systems, we remark that there is an assumption of familiarity with the four fundamental operations (addition, subtraction, multiplication, and division) in the base-10 number system. We take for granted an understanding of the terminology used in these processes (subtrahend and minuend, factors, divisor, dividend, etc.) and will refer to them without further comment.

3.2 Binary Number System

The binary number system is a fundamental concept in cybersecurity and computer science. It is the foundation upon which modern digital systems, including computers and electronic devices, are built. In this section, we will explore the basics of binary numbers and their properties.

3.2.1 Introduction to Binary Numbers

The binary number system, also known as the base-2 system, uses only two symbols, 0 and 1, to represent numerical values. That is to say, our familiar base-10 values (such as 47, 310, and 926) will be represented using a series of 0's and 1's. How this will be done is the notion upon which we will focus here.

The symbols 0 and 1 are called binary digits or bits. In contrast to the decimal system, which uses 10 digits (0–9), the binary system uses only two digits (0 and 1). The prefix "bi," which indicates "two," indicates the use of two symbols only.

The binary system is used in computer systems because electronic devices can easily distinguish between two states: on (represented by 1) and off (represented by 0). This simplicity makes it easier to design and build electronic circuits that can store and process binary data.

Example 3.3 Here are some examples of binary numbers:

0
1
10
11
100
101
110
111

When it is not clear by context which base is in use, binary numbers are traditionally followed by a subscripted "2," which indicates the base-value explicitly or by the word "binary" enclosed within parentheses. That is, a binary number would be presented either as 101001011_2 or as 101001011 (binary).

A worthy question at this moment is, "What decimal numbers are represented by these binary values?" It is to this question we now turn.

3.2.2 Binary Place Value and Positional Notation

Just as in the decimal system, each binary digit has a place value. However, in the binary system, the place values are powers of 2 instead of powers of 10. The rightmost bit has a place value of 2^0 (or 1), the next bit to the left has a place value of 2^1 (or 2), the next has a place value of 2^2 (or 4), and so on.

The equivalent base-10 value of a binary number is the sum of the products of each bit and its corresponding place value. Summation notation is particularly useful in the expression of numbers in base-2 notation (as well as other bases), but there is an alternative (somewhat more cumbersome) way we can determine the equivalent base-10 form of a binary number.

For bases other than 10, it is useful to begin our work with alternative number bases by constructing (or visualizing) a place-name grid as an aid to understanding the base-2 place names.

$2^7 = 128$	$2^6 = 64$	$2^5 = 32$	$2^4 = 16$	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$

The row beneath these place-name identifiers will contain 0's and 1's in each place-location. The decimal values of the digits in those place-name locations are found by multiplying 0 or 1 by the associated power of 2.

The ability to convert rapidly between base-2 and base-10 (and vice versa) is a skill that will improve your ability to perform mental calculations rapidly, and this is something we will find to be tremendously useful.

Example 3.4 Let's consider the binary number 1101. Each bit contributes to the total value as follows:

Using the place-name grid assists with understanding this breakdown:

$2^7 = 128$	$2^6 = 64$	$2^5 = 32$	$2^4 = 16$	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
				1	1	0	1

The value of the binary number 1101 is the sum of these contributions:

$8 + 4 + 0 + 1 = 13$ (in decimal form)

Example 3.5 Consider the binary number 10110. Using the formula above, we can calculate its decimal value as:

Or,

$2^7 = 128$	$2^6 = 64$	$2^5 = 32$	$2^4 = 16$	$2^3 = 8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$
			1	0	1	1	0

$16 + 0 + 4 + 2 + 0 = 22$

3.2.3 Converting Decimal Numbers to Binary

Converting decimal numbers to binary is an essential skill in discrete mathematics and computer science. There are two common methods for converting decimal numbers to binary: the division method and the subtraction method. Let's explore both methods with detailed examples.

3.2.3.1 Division Method

The division method involves repeatedly dividing the decimal number by 2 and keeping track of the remainders. The remainders, read from bottom to top, form the binary representation of the decimal number.

Here is the step-by-step process:

- Divide the decimal number by 2.
- Write down the remainder (0 or 1).
- Divide the quotient from step 1 by 2.
- Repeat steps 2 and 3 until the quotient becomes 0.

The binary representation is the sequence of remainders read from bottom to top.

Example 3.6 Convert the decimal number 75 to binary using the division method.

75 $75 \div \div 2 = 37$ remainder 1
$37 \div 2 = 18$ remainder 1
$18 \div \div 2 = 9$ remainder 0
$9 \div \div 2 = 4$ remainder 1
$4 \div \div 2 = 2$ remainder 0
$2 \div \div 2 = 1$ remainder 0
$1 \div \div 2 = 0$ remainder 1

Reading the remainders from bottom to top, we get the binary representation of 75:

75 (decimal) = 1001011 (binary).

Example 3.7 Convert the decimal number 2056 to binary using the division method.

$2056 \div 2 = 1028$ remainder 0
$1028 \div 2 = 514$ remainder 0
$514 \div 2 = 257$ remainder 0
$257 \div 2 = 128$ remainder 1
$128 \div 2 = 64$ remainder 0
$64 \div 2 = 32$ remainder 0
$32 \div 2 = 16$ remainder 0
$16 \div 2 = 8$ remainder 0
$8 \div 2 = 4$ remainder 0
$4 \div 2 = 2$ remainder 0
$2 \div 2 = 1$ remainder 0
$1 \div \div 2 = 0$ remainder 1

The binary representation of 2056, therefore, is 100000001000_2.

3.2.3.2 Subtraction Method

The subtraction method involves subtracting the largest possible power of 2 from the decimal number and repeating the process until the remainder is 0. The binary representation is formed by placing a 1 in the positions corresponding to the subtracted powers of 2 and a 0 in the remaining positions.

Here is the step-by-step process:

- Find the largest power of 2 that is less than or equal to the decimal number. The number of digits in the binary number will be the largest power, plus 1.
- Subtract the power of 2 from the decimal number.
- Repeat steps 1 and 2 with the remainder until the remainder becomes 0.
- Place a 1 in the positions corresponding to the subtracted powers of 2 and a 0 in the remaining positions.

Example 3.8 Convert the decimal number 75 to binary using the subtraction method.

The greatest power of 2 less than or equal to 75 is $2^6 = 64$. 75–64 (2^6) = 11, so place a 1 in the 2^6 position. Note the binary number will have 7 digits. Visualize this as a numerical representation having seven blank spaces: $\_\ \_\ \_\ \_\ \_\ \_\ \__2$. Remember, the "power position" is counted from the right, beginning at zero.

The greatest power of 2 that is less than or equal to 11 is $2^3 = 8$. This indicates to us the fifth and fourth power positions will be zeros; thus, we are at $1\ 0\ 0\ \_\ \_\ \_\ \__2$. We subtract 2^3 from 11:

$11-8 = 3$, so place a 1 in the third-power position, $1\ 0\ 0\ 1\ \_\ \_\ \__2$.

The greatest power of 2 that is less than or equal to 3 is $2^1 = 2$. Note that the second power position will be zero

$3-2 = 1$, so place a 1 in the 2^1 position to obtain: $1\ 0\ 0\ 1\ 0\ 1\ \__2$.

Finally, the greatest power of 2 less than or equal to 1 is 2^0, and $1-1 = 0$, so place a 1 in the 2^0 position.

The binary representation of 75 is:

75 (decimal) = 1001011_2 (binary).

The method used is a matter of personal preference; each will produce the same result. There is a benefit to the division method in that we automatically produce all the binary digits without having to keep track of the place names as we do in the subtraction approach.

3.2.4 Converting Binary Numbers to Decimal

We have already seen a method to convert a binary number to its decimal equivalent, using a table displaying the powers of 2. This method can be systematized using summation notation.

Assigning the binary digits values (from right to left) of $b_0, b_1, \ldots, b_{n-1}$, the transition to decimal form can be captured by:

$$\sum_{i=0}^{n-1} 2^i \cdot b_i$$

For instance, the binary number 101101_2 has coefficients $b_0 = 1$, $b_1 = 0$, $b_2 = 1$, and so on, up to and including $b_5 = 1$. Expansion of the summation notation yields:

$$1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 + 1 \cdot 2^5 = 1 + 0 + 4 + 8 + 0 + 32 = 45$$

Example 3.9 Convert 1011101101_2 to decimal form.

Observe that there are 10 digits in the binary value. It is common and convenient to subdivide binary numbers (due to their length) into groups of four digits, understanding that eight binary digits form a byte. Thus, our number would be more easily represented as $10\ 1110\ 1101_2$.

$$\begin{aligned} &1 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3 + 0 \cdot 2^4 + 1 \cdot 2^5 + 1 \cdot 2^6 + 1 \cdot 2^7 + 0 \cdot 2^8 + 1 \cdot 2^9 \\ &\quad = 1 + 0 + 4 + 8 + 0 + 32 + 64 + 128 + 0 + 256 = 749 \end{aligned}$$

If desired, the conversion can be performed working from left to right. We must count the number of binary digits present and recall the highest power of 2 will be one less than the number of binary digits.

Example 3.10 Convert $110\ 1011\ 0110_2$ to decimal form.

Solution: There are 11 digits in the binary number, so the highest power of 2 used will be $2^{10} = 1024$. We halve the powers of 2 as we proceed:

$$1 \cdot 2^{10} + 1 \cdot 2^9 + 0 \cdot 2^8 + 1 \cdot 2^7 + 0 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$$

$$= 1024 + 512 + 0 + 128 + 0 + 32 + 16 + 0 + 4 + 2 + 0$$

$$= 1718$$

3.2.5 Binary Arithmetic

Binary arithmetic involves performing mathematical operations on numbers in the binary system. The four basic arithmetic operations, recall, are addition, subtraction, multiplication, and division. Each of these operations in the binary system resembles the analogous process in the decimal system, but we must recognize the limitations on the value of digits being 0 or 1 only.

3.2.5.1 Binary Addition

To add two binary numbers, we add each column of bits, starting from the rightmost column. The result of this addition, by construction, must be 0, 1, or 2. If the sum in a column is 2, we write 0 and carry 1 to the next column. If the sum is 3, we write 1 and carry 1 to the next column.

Example 3.11 Add 1011 and 1101 in binary.

	1	0	1	1
	1	1	0	1
1	1	0	0	0

3.2.5.2 Binary Subtraction

To subtract two binary numbers, we subtract each column of bits, starting from the rightmost column. If the top bit is smaller than the bottom bit, we borrow 1 from the next column to the left. A few words must be said about the "borrowing" process in binary, because our work in base-10 is so familiar to us that we often overlook what we mean by "borrow."

Base-10 subtraction involving borrowing looks like this:

	3	6	9	4
–		8	6	5

Since 5 exceeds 4, we must borrow from the column to the left; typically, we place a "slash" through the 9 (the number from which we are borrowing), write an 8 above it, and place a 1 in front of the 4:

	3	6	8 ~~9~~	14
–		8	6	5

For the first time, perhaps, we must be careful regarding what this "1" means…we are borrowing from the place to the left of the four (in this case, the tens place), and so we are "borrowing" 10. This is what transforms the "4" into "14," so subtraction may proceed.

	3	6	8 ~~9~~	14
–		8	6	5
			2	9

Note, we could borrow from the next place to the left, but in base-10 we can simply subtract 8 from 36, to obtain:

	3	6	8 ~~9~~	14
–		8	6	5
	2	8	2	9

In binary, when we "borrow" from the place to the left, we must remember we are borrowing a power of 2. This is a challenging concept to understand. As an aid, let us present the following rule set for the subtraction of binary numbers:

$0 - 0 = 0$

$1 - 1 = 0$

$1 - 0 = 1$

$0 - 1 = 1$, borrowing "1" from the column to the left

All but the last of these rules appear elementary; the final rule seems nonsensical (because we are used to working with decimal numbers). Why would this be so?

Let us begin with the subtraction of the following binary numbers:

10_2
$\underline{1_2}$

Our first subtraction is 0−1, and so we must borrow. When we borrow from the position to the left, our 0 changes to a 2, and so we have:

02_2
$-\underline{1_2}$

From a technical standpoint, writing the digit "2" in a binary number is improper, but it serves to illustrate our point here. When we subtract, 2−1 = 1, and so our result is:

10_2
$-\underline{1_2}$

Our first subtraction is 0–1, and so we must borrow. When we borrow from the position to the left, our 0 changes to a 2, and so we have:

0^20_2
$-\underline{1_2}$
1_2

Example 3.12 Subtract 1011 from 11101 in binary.

```
11101
−1011
------------
10010
```

The first subtraction (starting from the right) results in 0. The second subtraction, however, involves borrowing. Considering our rule for subtraction:

0 – 1 = 1, Borrowing "1" from the column to the left.

Our second subtraction results in "1," with the value in the column to the left being reduced from 1 to 0.

One can always check the accuracy of binary subtraction by converting all values to decimal form and comparing the results there:

$$11101 = 1 + 4 + 8 + 16 = 29$$

$$1011 = 1 + 2 + 8 = 11$$

Difference is 18

Binomial representation of 18 is found via:

18/2 = 9 Remainder 0
9/2 = 4 Remainder 1
4/2 = 2 Remainder 0
2/2 = 1 Remainder 0
1/2 = 0 Remainder 1

The binomial representation of the result is 10010_2.

Some individuals feel (particularly at the outset) it is more accurate to perform the operations in this way. That is, binary numbers are converted to decimal numbers, the operation of the problem is performed in base-10, and the result is converted back to binary. There is nothing fundamentally wrong with this approach, but it is cumbersome.

A favored expression of our authors is that we can almost always create workarounds for processes we are attempting. The trouble lies in the fact that workarounds often create additional steps that proliferate the possibility of errors, and in that, they forestall your acquisition of knowledge regarding new concepts.

3.2.5.3 Binary Multiplication

To multiply two binary numbers, we multiply each bit of the multiplicand (top number) by the multiplier (bottom number) and add the intermediate products.

Example 3.13 Multiply 1011 by 101 in binary.

		1	0	1	1
	×		1	0	1
		1	0	1	1
	0	0	0	0	
1	0	1	1		
1	1	0	1	1	1

Again, observe how conversion to base-10 makes clear the correctness of the result:

$$1011_2 = 8 + 2 + 1 = 11$$

$$101_2 = 4 + 1 = 5$$

In base 10, the product is 55; the binary product was:

$$110111_2 = 32 + 16 + 4 + 2 + 1 = 55$$

Observe that, while the conversion to base-10 form is useful as a check, it ultimately becomes tedious. It is far better to become proficient in the calculations within base-2, making this conversion/comparison redundant.

3.2.5.4 Binary Division

To divide two binary numbers, we use the long division algorithm, like decimal division. We divide the dividend (top number) by the divisor (bottom number), writing the quotient above the division bracket and the remainder as the result of the final subtraction in the long division process.

If the dividend is smaller than the divisor, the division results in a proper fraction equal to the dividend over the divisor.

If the divisor is smaller than the dividend, we must perform the long division process to find the quotient and remainder.

A challenge exists in recognizing which of two binary numbers is larger than the other. Trivially, a binary number having fewer digits is always smaller than a binary number having more digits.

If the two binary numbers have the same number of digits, the first number having a zero (left to right) is the smaller of the two numbers.

Example 3.14 Divide 100110 by 1011 in binary.

The dividend is larger than the divisor, since the dividend has the greater number of binary digits. Therefore, we must perform long division.

Having decided the dividend is greater than the divisor, we recall the principles of division in base-10. Since the divisor has four digits, we look to the dividend and notice the binary number consisting of the first four digits of the dividend is smaller than the divisor; therefore, we consider the binary number represented by the first five digits of the dividend. That five-digit binary number is trivially larger than the divisor, so the divisor must divide into that number by a multiple of 1_2.

Using the notation and technique of base-10 long division and recalling our rules for base-2 subtraction, we find:

```
                          1
1   0   1 | 1   0   0   1   1   0
          |     1   0   1   1
          |-----------------------
          |     1   0   0   0   0
```

By construction, the four-digit binary number resulting from that subtraction is smaller than the divisor, so we "bring down" the next digit in the dividend.

Now, we iterate the process and consider if 1011_2 is less than 10000_2, which it is (comparison of the number of digits makes this clear). We multiply 1011_2 by 1 to obtain:

```
                          1   1
1   0   1 | 1   0   0   1   1   0
          |     1   0   1   1
          |-----------------------
          |     1   0   0   0   0
          |         1   0   1   1
          |-----------------------
          |         0   0   0   1
```

1011_2 is not less than 1_2, and thus we have reached the remainder. So, in binary form, $100110 \div 1011 = 11$ with a remainder of 1.

Example 3.15 Divide 1101011 by 101 in binary.

We begin by observing dividend possesses more digits than the divisor, and so it is the larger of the two binary numbers.

Next, we compare the first three digits of the dividend with the divisor and find 110 to be larger than 101; therefore,

```
                      1
1   0   1 | 1   1   0   1   0   1   1
          |-1   0   1
          |-----------
              1
```

We bring down the next digit in the dividend and find the binary digit 11 to be smaller than the divisor. A zero is placed in the quotient, and the next digit of the dividend is brought down.

					1	0			
1	0	1	1	1	0	1	0	1	1
			−1	0	1				
					1	1	0		

We see 110 is larger than 101, so the next digit in the quotient is a 1.

					1	0	1		
1	0	1	1	1	0	1	0	1	1
			−1	0	1				
					1	1	0		
					−1	0	1		
							1		

We bring down the next digit of the dividend and find again that 11 is smaller than the divisor. A zero is placed in the quotient, the next digit of the dividend is brought down, and the process continues:

					1	0	1	0	1
1	0	1	1	1	0	1	0	1	1
			−1	0	1				
					1	1	0		
					−1	0	1		
							1	1	1
							−1	0	1
								1	0

There are no more digits in the dividend to be brought down, and the final difference is the remainder. Therefore, our quotient (in binary) is 10101, and the remainder is 10.

As an exercise, we ask you to show the dividend (in decimal form) is 107, while the divisor is 5. The quotient is 21, and the remainder is 2. We ask you to show the binary equivalents of the long division shown above match the decimal result.

3.2.6 Signed Binary Numbers

We have, thus far, discussed unsigned binary numbers, which represent non-negative integers. However, in many applications, we need to represent both positive and negative numbers. To do so, we require the use of signed binary numbers.

There are two common methods for representing signed binary numbers: sign-magnitude representation and two's complement representation.

3.2.6.1 Sign-Magnitude Representation

Definition Most significant bit (MSB): The leftmost bit in any number's binary representation.

Definition *n*-bit representation of a binary number: A representation of a binary number required to show precisely n bits (individual digits in the representation).

In sign-magnitude representation, the most significant bit is used to indicate the sign of the number. A 0 in the MSB indicates the number is positive, while a 1 in the MSB indicates the number is negative. The remaining bits represent the magnitude (unsigned) of the number.

Example 3.16 In the following 8-bit sign-magnitude representations, identify if the number is positive or negative, and then state its value in base-10.

a) 0100 1011
b) 1101 0001
c) 1000 0101

Solution:

a) The MSB is 0, indicating the number is positive. The remaining digits, 1001011, represent the decimal number 75. The number represented by the sign-magnitude format is 75.
b) The MSB is 1, indicating the number is negative. The remaining digits, 1010001, represent the decimal number 81. The number represented by the sign-magnitude format is −81.
c) The MSB is 1, indicating the number is negative. The remaining digits, 0000101, represent the decimal number 5. The number represented by the sign-magnitude format is −5.

Example 3.17 Represent the following decimal numbers in 8-bit sign-magnitude representation.

a) 5
b) −22
c) −101

Solution:

a) The representation of positive numbers utilizes a 0 in the MSB, so the sign-magnitude representation of the +5 begins with 0. The decimal value 5 (in binary) is 101, and (fleshing out the 8-bit representation with zeros) so the sign-magnitude representation of +5 is 0000 0101.
b) Since the number is negative, the sign-magnitude representation will begin with 1. The binary equivalent of 22 is 10110, so the 8-bit sign-magnitude representation of −22 will be 1001 0110.
c) As seen in example b, the MSB will be 1. The binary equivalent of 101 is 1100100, and so the 8-bit sign-magnitude representation of −101 is 1100 0100.

Example 3.18 What are the decimal representations of the two 8-bit sign-magnitude numbers having the greatest absolute value?

Solution: The two values will begin with either a 0 or a 1 (identifying the sign of the number), and to achieve the greatest possible absolute value, all of the binary digits will be 1's. There are seven remaining bits, and so the problem amounts to finding the decimal equivalent of 1111111. This is $64 + 32 + 16 + 8 + 4 + 2 + 1 = 127$, and thus the two decimal values having the greatest possible absolute value which we can represent in 8-bit sign-magnitude form are −127 and 127.

One (possibly non-obvious) drawback of sign-magnitude representation is that our definition ultimately leads to two representations for zero. In 8-bit sign-magnitude representation, both +0

(0000 0000) and −0 (1000 0000) can appear. This can lead to complications in arithmetic operations, which we can overcome through the inclusion of a test for equality of two binary numbers. At first, it may seem that there is no serious issue to debate here, but it is something that arises in computations, and about which we must be aware.

3.2.6.2 Two's Complement Representation

Two's complement representation is the most widely used method for representing signed binary numbers in modern computers. In two's complement, positive numbers are represented in the same way as unsigned binary numbers; the distinction lies in how negative numbers are represented.

For a negative binary number, first find the binary representation of the absolute value of the number. Next, invert all the binary digits (changing 0's to 1's and vice versa) and add 1 to the result.

Example 3.19 Represent both +5 and −5 in 8-bit two's complement representation.

+5 = 0000 0101

To represent −5, we invert the bits in the binary representation of 5 and then add 1.

−5 = 1000 0101

Invert the bits:

0111 1010

Add 1:

0111 1011

Therefore, −5 = 0111 1011 in two's complement representation.

In two's complement representation, the range of numbers that can be represented with n bits is from −(2^(n − 1)) to +(2^(n − 1) − 1). For example, with 8 bits, we can represent numbers from −128 to +127.

One advantage of two's complement representation is that it simplifies arithmetic operations. Addition and subtraction can be performed in the same way as with unsigned binary numbers, and the result will be correct, even if it involves negative numbers.

Example 3.20 Add +5 and −3 in 8-bit two's complement representation.

Solution: We have seen +5 = 0000 0101, it remains to obtain the representation of −3. The representation of 3 in 8-bit format is 0000 0011; we invert the bits to obtain 1111 1100, then add 1 to the result, obtaining 1111 1101; at this point, we add:

+5 = 0000 0101
−3 = 1111 1101

0000 0010.

Note that, in base-10, the result is +2.

Observe the "carry" from the addition in the left-most bit is suppressed. That is, if we were adding two base-2 numbers not employing this technique, the result would have been 1 0000 0000;

this produces 9 bits, violating the precept we are using 8-bit representation. We adopt the procedure of truncating the digits at the point of the 8th bit.

Example 3.21 Add −12 and −3 in 8-bit two's complement representation.

The 8-bit two's complement representation of −12 can be obtained using the conversion method described earlier:

The base-2 representation of 12 is 0000 1100, from which we invert the bits to produce 1111 0011; we add 1 to this value and find 1111 0100. From the prior example, the two's complement form of −3 is 1111 1101.

Next, we add:

```
1111 0100
+1111 1101
----------
1111 0001
```

The initial 1 indicates this is a negative number, but what number IS it? We know the result should be −15 in decimal form, but is it? We convert the remaining digits to two's complement and see what we obtain:

```
0001110V
+       1
```

0001111 = 1 + 2 + 4 + 8 = 15, changed to negative due to the original 1 from our sum, so we obtain −15.

3.3 Octal (Base-8) Number System

In this section, we will explore the octal number system, also known as the base-8 numbering system, which is another positional number system used in computer science. While not as widely used as the binary or hexadecimal systems, understanding octal numbers can provide insights into the relationships between different number systems and their applications in computer science.

In computing environments, the octal number system is commonly used as a shorter representation of binary numbers by grouping binary digits into threes. The **chmod** command in Linux or UNIX uses octal to assign file permissions.

3.3.1 Introduction to Octal Numbers

The octal number system uses eight symbols to represent numbers: 0, 1, 2, 3, 4, 5, 6, and 7.

Octal numbers are often used in computing as a compact representation of binary numbers. Each octal digit corresponds to three binary digits (bits), making it more convenient for humans to read and write than long strings of binary digits.

As before, we can indicate (unless the context is clear) these to be octal numbers by including the word "octal" in parentheses to the right of the number, or with a subscripted "8," or by specifically stating we are working with octal numbers.

3.3.2 Converting Decimal to Octal

To convert a decimal number to its octal equivalent, we use the division method introduced for converting decimal to binary. The main difference is that instead of dividing by 2, we divide by 8 and keep track of the remainders, continuing until the quotient is zero.

Here's the step-by-step process:

- Divide the decimal number by 8.
- Write down the remainder (0 to 7).
- Divide the quotient from step 1 by 8.
- Repeat steps 2 and 3 until the quotient becomes 0.

The octal representation is the sequence of remainders read from bottom to top.

Example 3.22 Convert the decimal number 492 to octal.

Solution:

$492 \div 8 = 61$ remainder 4
$61 \div 8 = 7$ remainder 5
$7 \div 8 = 0$ remainder 7
Reading the remainders from bottom to top, we get the octal representation of 492:

492 (decimal) = 754 (octal).

Example 3.23 Convert the decimal number 1023 to octal.

Solution: $1023 \div 8 = 12$
7 remainder $7127 \div 8 = 15$ remainder 7
$15 \div 8 = 1$ remainder 7
$1 \div 8 = 0$ remainder 1
Reading the remainders from bottom to top, we get the octal representation of 1023:

1023 (decimal) = 1777 (octal).

3.3.3 Converting Octal to Decimal

Like the other positional number systems, each octal digit has a place value. The place values are powers of 8, increasing from right to left. The rightmost digit has a place value of 8^0 (or 1), the next digit to the left has a place value of 8^1 (or 8), the next has a place value of 8^2 (or 64), and so on.

The value of an octal number is the sum of the products of each digit and its corresponding place value. Using summation notation:

$$\sum_{i=0}^{n-1} 8^i \cdot b_i$$

Example 3.24 Confirm the result of the conversion of the decimal number 492 to 754_8 by converting the octal number to decimal form.

Solution:

$$\sum_{i=0}^{2} 8^i \cdot b_i = 1 \cdot 4 + 8 \cdot 5 + 7 \cdot 64 = 4 + 40 + 448 = 492$$

Example 3.25 Confirm the conversion of the decimal number 1023 to 1777_8 by converting the octal number to decimal form.

Solution:

$$\sum_{i=0}^{3} 8^i \cdot b_i = 1 \cdot 7 + 8 \cdot 7 + 7 \cdot 64 + 1 \cdot 512 = 7 + 56 + 448 + 512 = 1023$$

Example 3.26 Convert the octal number 357 to decimal form.

Solution:

$$\sum_{i=0}^{2} 8^i \cdot b_i$$

$$7 \cdot 8^0 + 5 \cdot 8^1 + 3 \cdot 8^2 = 7 + 40 + 192 = 239$$

Example 3.27 Convert the octal number 2045 to decimal form.

Solution:

$$\sum_{i=0}^{3} 8^i \cdot b_i$$

$$5 \cdot 8^0 + 4 \cdot 8^1 + 0 \cdot 8^2 + 2 \cdot 8^3 = 5 + 32 + 192 = 1061$$

3.3.4 Converting Binary to Octal

Because we have a method for converting binary numbers to decimal numbers, and a method for converting decimal numbers to octal numbers, we already know we can convert binary numbers to octal form through an intermediate step into decimal form. There is a means by which we can make the conversion directly, cutting the base-10 format (which we do not need in this problem!) from the process.

The idea behind converting binary numbers to octal stems from the realization each octal digit corresponds to three binary digits (bits).

What do we mean by this? Consider the following equivalent decimal, binary, and octal numbers (showing the binary number form with three digits):

Decimal	Binary	Octal
0	000	0
1	001	1
2	010	2
3	011	3
4	100	4
5	101	5
6	110	6
7	111	7

To convert a binary number to octal, we simply group the binary digits in sets of three, starting from the rightmost digit and moving left. If the leftmost group has fewer than three digits, we pad it with leading zeros to complete the group.

That is,

- Group the binary digits in sets of three, starting from the rightmost digit and moving left.
- If the leftmost group has fewer than three digits, pad it with leading zeros.
- Convert each group of three binary digits to its octal equivalent.
- Concatenate the octal digits to obtain the octal number.

Example 3.28 Convert the binary number 101101001 to octal.

Solution:

Step 1: Group the binary digits in sets of three from right to left.

101 101 001

Step 2: The leftmost group has three digits, so no padding is needed.

Step 3: Convert each group of three binary digits to its octal equivalent.

101 = 5

101 = 5

001 = 1

Step 4: Concatenate (place side-by-side, in sequence) the octal digits.

101 101 001 (binary) = 551 (octal)

Example 3.29 Convert the binary number 1011001110 to octal.

Solution:

Step 1: Group the binary digits in sets of three from right to left.

1 011 001 110

Step 2: Pad the leftmost group with two leading zeros.

001 011 001 110

Step 3: Convert each group of three binary digits to its octal equivalent.

001 = 1

011 = 3

001 = 1

110 = 6

Step 4: Concatenate the octal digits.

1011001110 (binary) = 1316 (octal)

3.3.5 Converting Octal to Binary

Converting octal numbers to binary is the reverse process of converting binary to octal. Since each octal digit corresponds to three binary digits (bits), we can convert an octal number to binary by replacing each octal digit with its equivalent three-bit binary representation.

Here's the step-by-step process:

- Write down the three-bit binary equivalent for each octal digit.
- Concatenate the binary representations to obtain the binary number.

Example 3.30 Convert the octal number 735 to binary.

Solution:

Step 1: Write down the three-bit binary equivalent for each octal digit.

7 = 111

3 = 011

5 = 101

Step 2: Concatenate the binary representations.

735 (octal) = 111011101 (binary)

Example 3.31 Convert the octal number 2047 to binary.

Solution:

Step 1: Write down the three-bit binary equivalent for each octal digit.

2 = 010

0 = 000

4 = 100

7 = 111

Step 2: Concatenate the binary representations.

2047 (octal) = 010000100111 (binary)

When converting octal numbers to binary, the resulting binary number will always have a number of digits that is a multiple of three (provided we show any padded-out zeros at the front of the numerical representation). This is because each octal digit is represented by exactly three binary digits. Leading zeros can, if desired, be omitted from the binary representation.

3.3.6 Addition and Subtraction in Octal

When adding or subtracting octal numbers, we follow the same rules as in decimal arithmetic, but with a smaller set of digits.

Addition in Octal:

To add two octal numbers, we add the digits in each column, starting from the rightmost column. If the sum in any column is 8 or greater, we carry over the excess to the next column to the left. Because of the growing complexity, it is useful to create a carry row above the octal numbers being added.

Example 3.32 Adding two octal numbers

3 7 2 (base 8) + 5 4 6 (base 8)

Solution:

1	1	1		Carry Row
	3	7	2	
+	5	4	6	
1	1	4	0	

The sum is 1048 (base 8).

Subtraction in Octal:

To subtract two octal numbers, we subtract the digits in each column, starting from the rightmost column. If the top digit in any column is smaller than the bottom digit, we borrow from the next column to the left.

Example 3.33 Subtracting two octal numbers

7 5 3 (base 8) – 2 6 4 (base 8)

Solution:

	7	5	3
–	2	6	4

Since 4 exceeds 3, we must borrow from the 8's place before subtracting the ones-place digits:

	7	4	11
–	2	6	4
			7

Again, 6 exceeds 4, so we must borrow from the 64's place before subtracting the 8's place digits:

	6	12	11
–	2	6	4
		6	7

Finally, we can subtract 2 from 6 in the 64's place:

	6	12	11
–	2	6	4
	4	6	7

The difference is 467 (base 8).

As a check, we can convert to base-10 and perform the subtraction. The equivalences are $753_8 = 491$ and $264_8 = 180$; the difference of the base-10 numbers is 311, which is precisely the decimal equivalent of 467_8.

When performing addition and subtraction in octal, it's essential to be mindful of the base and to ensure that all digits are within the valid range (0–7). Practice and familiarity with the octal number system will help you perform these operations more efficiently.

Example 3.34 Calculate $625_8 + 374_8 - 743_8$. Check the result through conversion to base-10.

Solution:

1	1	1		Carry Row
	6	2	5	
+	3	7	4	
1	2	2	1	

1	2	2	1
–	7	4	3

For the first subtraction, we must borrow from the 8's place:

1	2	1	9
–	7	4	3
			6

For the next subtraction, we must borrow from the 64's place:

1	1	9	9
–	7	4	3
		5	6

For the next (and final) subtraction, we must borrow from the 512's place:

0	9	9	9
–	7	4	3
	2	5	6

The result is 258_8.

To test our computations, we note $625_8 = 405$, $374_8 = 252$, $743_8 = 483$, and $256_8 = 174$. Computation verifies:

$$405 + 252 - 483 = 174$$

3.3.7 Multiplication and Division in Octal

Multiplication and division in the octal number system follow similar principles to those in the decimal system, but with a different set of rules for carrying and handling remainders.

When multiplying two octal numbers, we multiply each digit of the first number by each digit of the second number, starting from the right-most digits. We then add the intermediate products, aligning them based on their place values and carrying when necessary.

To familiarize you with the notation we will employ, we will first perform a multiplication in base-10.

Example 3.35 Multiply the decimal numbers 2518 and 27.

Solution:

					Carry Row for Multiplication by 7
	2	5	1	8	
×			2	7	
	3	1	5		Carry Row for Multiplication by 7
	2	5	1	8	
×			2	7	
1	7	6	2	6	
	1		1		Carry Row for Multiplication by 2
	2	5	1	8	
×			2	7	
1	7	6	2	6	
5	0	3	6		
	1		1		Carry Row for Multiplication by 2
	2	5	1	8	
×			2	7	
1	7	6	2	6	
5	0	3	6		
6	7	9	8	6	Sum of the multiplication rows

The product is 67,986.

Before multiplying larger octal numbers, our first illustration will be with single-digit values. If we multiply $3_8 \times 5_8$, what will we obtain? Since 3 and 5 are both less than 8, their decimal representations are also 3 and 5, so the decimal product is 15. In base-8, this is 17_8.

Observe how we would perform this calculation using the common multiplication notation:

1		Carry Row
	3	
×	5	
1	7	

Next, consider $23_8 \times 5_8$.

	1		Carry Row
	2	3	
	×	5	
		7	

1	1		Carry Row
	2	3	
	×	5	
	3	7	

1	1		Carry Row
	2	3	
	×	5	
1	3	7	

Converting to base-10, we have $23_8 = 19$, and $19 \times 5 = 95$. The product $137_8 = 95$, as well.

Example 3.36 Multiply 235 (base 8) × 42 (base 8)

Solution:

			Carry Row
	2	3	5
	×	4	2

$5 \times 2 = 10$; this is equivalent to 12_8, so we write a 2 below the right-most column, and carry a 1.

		1		Carry Row
	2	3	5	
	×	4	2	
			2	

Next, we multiply $2 \times 3 = 6$, and add the 1 that had been carried, to give 7. Since this is a suitable octal digit, we write 7 in the next column to the left and perform no carry.

		1		Carry Row
	2	3	5	
	×	4	2	
		7	2	

Next, we multiply $2 \times 2 = 4$. Since this is a suitable octal digit, we write 4 in the next column and perform no carry.

			Carry Row
	1		
2	3	5	
×	4	2	
4	7	2	

Now we clear the carry row and repeat the process, multiplying by $4.4 \times 5 = 20$, which is 24_8. We write a 4 beneath the 7 and carry the 2.

			Carry Row
	2		
2	3	5	
×	4	2	
4	7	2	
	4		

Next, $4 \times 3 = 12$, to which we add the carried 2 to obtain 14. In base-8, this is 16, so we write a 6 in the next column to the left and carry the 1.

				Carry Row
	1	2		
	2	3	5	
	×	4	2	
	4	7	2	
	6	4		

Finally, we multiply $4 \times 2 = 8$, to which we add the carried 1 to give 9. In base-8, this is 11. We write 1 in the next column to the left and carry the 1.

				Carry Row
1	1	2		
	2	3	5	
	×	4	2	
	4	7	2	
1	6	4		

$4 \times 0 = 0$, to which we add the carried 1, which we write in the next column to the left.

					Carry Row
	1	1	2		
		2	3	5	
		×	4	2	
		4	7	2	
1	1	6	4		

We add the two product rows using base-8 addition,

	1	1	2		Carry Row
		2	3	5	
		×	4	2	
		4	7	2	
1	1	6	4		
1	2	3	3	2	

The final product is 12332_8.

Observe $235_8 = 157$ and $42_8 = 34$. The product of these values is 5338, which is 12332_8.

Division in Octal:

To divide two octal numbers, we follow a process similar to long division in the decimal system. We divide the left-most digit(s) of the dividend by the divisor, write down the quotient, multiply the quotient by the divisor, subtract the result from the dividend, and bring down the next digit of the dividend. We repeat this process until we have processed all the digits of the dividend.

Example 3.37 Divide the two octal numbers $6354_8 \div 43_8$. Express the result to the nearest octal digit.

Solution: Set up the division as you would in base-10:

4	3	6	3	5	4

We note that 43 does not divide into 6 but does divide 1 time into 63:

			1		
4	3	6	3	5	4
		−4	3		
		2	0		

We bring down the next digit and repeat the process:

			1	3	
4	3	6	3	5	4
		−4	3		
		2	0	5	
		−1	5	1	
			3	4	

Note: It is important to realize we are dividing base-8. If we multiply 43_8 by 3_8, the result is 151_8. You should convince yourself of this.

			1	3	
4	3	6	3	5	4
		−4	3		
		2	0	5	
		−1	5	1	
			3	4	4

			1	3	6
4	3	6	3	5	4
		−4	3		
		2	0	5	
		−1	5	1	
			3	4	4
			−3	2	2
				2	2

			1	3	6	0.4	0
4	3	6	3	5	4	0.0	0
		−4	3				
		2	0	5			
		−1	5	1			
			3	4	4		
			−3	2	2		
				2	2	0	
				−2	1	4	
						4	0

The result, rounded to one octal place, is 136.4_8.

3.4 Hexadecimal (Hex) Number System

In this section, we will explore the hexadecimal (hereafter, also referred to as "hex," or base-16) number system, which is widely used in computer science and digital systems. Hex numbers provide a compact and convenient way to represent large binary numbers, making them particularly useful for expressing memory addresses, color codes, and other values in computing.

3.4.1 Introduction to Hexadecimal Numbers

The hexadecimal number system uses 16 distinct symbols to represent numbers. These symbols include the digits 0 through 9 and the letters A through F, where A represents 10, B represents 11, C represents 12, D represents 13, E represents 14, and F represents 15.

Here are the hexadecimal digits and their decimal equivalents:

Hexadecimal	Decimal
0	0
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
A	10
B	11
C	12
D	13
E	14
F	15

Here are some examples of hexadecimal numbers:

7
1F
3B
A9
FF
2C4

Like other positional number systems, each hexadecimal digit has a place value. The place values are powers of 16, increasing from right to left.

That is, the rightmost digit has a place value of 16^0 (or 1), the digit to its left has a place value of 16^1 (or 16), the next has a place value of 16^2 (or 256), and so on.

3.4.2 Converting Decimal to Hexadecimal

To convert a decimal number to its hexadecimal equivalent, we can use a method like the division method used for converting decimal to binary and octal. The main difference is that instead of dividing by 2 or 8, we divide by 16 and keep track of the remainders.

Here is the process:

- Divide the decimal number by 16.
- Write down the remainder (0 to 15) as a hexadecimal digit (0–9, A–F).
- Divide the quotient from step 1 by 16.
- Repeat steps 2 and 3 until the quotient becomes 0.
- The hexadecimal representation is the sequence of remainders read from bottom to top.

Example 3.38 Convert the decimal number 2748 to hexadecimal.

Solution: $2748 \div 16 = 171$ remainder 12 (C)
$171 \div 16 = 10$ remainder 11 (B)
$10 \div 16 = 0$ remainder 10 (A)
Reading the remainders from bottom to top, we get the hexadecimal representation of 2748:

2748 (decimal) = ABC (hexadecimal).

Example 3.39 Convert the decimal number 31459 to hexadecimal.

Solution:

$31459 \div 16 = 1966$ remainder 3
$1966 \div 16 = 122$ remainder 14 (E)
$122 \div 16 = 7$ remainder 10 (A)
$7 \div 16 = 0$ remainder 7

Reading the remainders from bottom to top, we get the hexadecimal representation of 31459:31459 (decimal) = 7AE3 (hexadecimal).

It is useful to recognize that when converting decimal to hexadecimal representations, the hex form will always have fewer, or an equal number of digits compared to its decimal form. This is because each hexadecimal digit can represent a larger range of values (0 to 15) than a decimal digit (0 to 9).

3.4.3 Converting Hexadecimal to Decimal

The decimal value of a hexadecimal number is the sum of the products of each digit and its corresponding place value (a power of 16). That is, the value of a hex number is given by:

$$\sum_{i=0}^{n-1} 16^i \cdot b_i$$

Example 3.40 Convert the hexadecimal number 1A5 to decimal form.

Solution: Each digit contributes to the total value, using the summation formula for conversion from hex to decimal form:

$$\sum_{i=0}^{2} 16^i \cdot b_i 5 \cdot 16^0 + 10 \cdot 16^1 + 1 \cdot 16^2 = 5 + 160 + 256 = 421$$

Example 3.41 Convert the hexadecimal number 2CF to decimal form.

Solution: Using the summation formula for conversion from hex to decimal form:

$$\sum_{i=0}^{2} 16^i \cdot b_i = 15 \cdot 16^0 + 12 \cdot 16^1 + 2 \cdot 16^2 = 15 + 192 + 512 = 719$$

Example 3.42 Convert the hexadecimal number 1D3 to decimal form.

Solution:

$$\sum_{i=0}^{2} 16^i \cdot b_i = 3 \cdot 16^0 + 13 \cdot 16^1 + 1 \cdot 16^2 = 3 + 208 + 256 = 467$$

Converting hexadecimal numbers to decimal is useful when working with computer systems, as hexadecimal is often used to represent various types of data, such as memory addresses and color codes.

3.4.4 Converting Binary to Hexadecimal

As we saw in the conversion from binary to octal, converting binary numbers to hex is a relatively straightforward process because each hexadecimal digit corresponds to four binary digits (bits).

That is:

Decimal	Binary	Hex
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

To convert a binary number to hexadecimal, we group the binary digits in sets of four, starting from the rightmost digit and moving left. If the leftmost group has fewer than four digits, we pad it with leading zeros to complete the group.

Here is the step-by-step process:

- Group the binary digits in sets of four, starting from the rightmost digit and moving left.
- If the leftmost group has fewer than four digits, pad it with leading zeros.
- Convert each group of four binary digits to its hexadecimal equivalent.
- Concatenate the hexadecimal digits to obtain the hexadecimal number.

Example 3.43 Convert the binary number 1101011001 to hexadecimal.

Solution:

Step 1: Group the binary digits in sets of four from right to left.

0110 1011 0010

Step 2: The leftmost group has four digits, so no padding is needed.

Step 3: Convert each group of four binary digits to its hexadecimal equivalent.

$0110 = 6$

$1011 = \mathrm{B}$

$0010 = 2$

Step 4: Concatenate the hexadecimal digits.

$1101011001\ (\text{binary}) = 2B6\ (\text{hexadecimal})$

Example 3.44 Convert the binary number 1011101110101 to hexadecimal.

Solution:

Step 1: Group the binary digits in sets of four from right to left.

10111 0111 0101

Step 2: Pad the leftmost group with three leading zeros.

0001 0111 0111 0101

Step 3: Convert each group of four binary digits to its hexadecimal equivalent.

$0001 = 1$

$0111 = 7$

$0111 = 7$

$0101 = 5$

Step 4: Concatenate the hexadecimal digits.

$1011101110101\ (\text{binary}) = 1775\ (\text{hexadecimal})$

3.4.5 Converting Hexadecimal to Binary

Converting hexadecimal numbers to binary is the reverse process of converting binary to hexadecimal. Since each hexadecimal digit corresponds to four binary digits (bits), we can convert a hexadecimal number to binary by replacing each hexadecimal digit with its equivalent four-bit binary representation.

That is:

- Write down the four-bit binary equivalent for each hexadecimal digit.
- Concatenate the binary representations to obtain the binary number.

Example 3.45 Convert the hexadecimal number 7E3 to binary.

Solution:

Recall:

Decimal	Binary	Hex
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

Step 1: Write down the four-bit binary equivalent for each hexadecimal digit.

7 = 0111

E = 1110

3 = 0011

Step 2: Concatenate the binary representations.

7E3 (hexadecimal) = 011111100011 (binary)

Example 3.46 Convert the hexadecimal number 2A9F to binary.

Solution:

Step 1: Write down the four-bit binary equivalent for each hexadecimal digit.

2 = 0010

A = 1010

9 = 1001

F = 1111

Step 2: Concatenate the binary representations.

2A9F (hexadecimal) = 0010101010011111 (binary)

When converting hexadecimal numbers to binary, the resulting binary number will always have a number of digits that is a multiple of four (if padded zeros are counted). This is because each hexadecimal digit is represented by exactly four binary digits.

3.5 Ternary Number System

The ternary number system, also known as the base-3 system, uses three distinct symbols to represent numbers: 0, 1, and 2. In the ternary system, each place-value position corresponds to a power of 3, increasing from right to left.

The place values in the ternary system are as follows:

...	3^3	3^2	3^1	3^0
...	27	9	3	1
...	(0–2)	(0–2)	(0–2)	(0–2)

To determine the decimal value of a ternary number, we multiply each ternary digit by its corresponding place value and then sum up the results.

In summation notation,

$$\sum_{i=0}^{n-1} 3^i \cdot b_i$$

Example 3.47 Convert the ternary number 1201_3 to decimal.

Step 1: Multiply each digit by its corresponding place value.

$$\begin{aligned}\sum_{i=0}^{3} 3^i \cdot b_i &= 1 \cdot 3^0 + 0 \cdot 3^1 + 2 \cdot 3^2 + 1 \cdot 3^3 \\ &= 1 + 0 + 18 + 27 \\ &= 46\end{aligned}$$

Therefore, the decimal value of the ternary number 1201 is 46.

To convert a decimal number to its ternary equivalent, we can use a method like the division method used for converting decimal to binary, octal, or hexadecimal. The difference is that we divide by 3, since this is the base of the ternary system.

Example 3.48 Convert the decimal number 75 to ternary.

$75 \div 3 = 25$ remainder 0
$25 \div 3 = 8$ remainder 1
$8 \div 3 = 2$ remainder 2
$2 \div 3 = 0$ remainder 2

Reading the remainders from bottom to top, we get the ternary representation of 75:
75 (decimal) = 2210 (ternary).

The ternary number system has been used in computer systems such as the Russian Setun computer in the 1950s and in certain digital logic designs. Ternary logic has also been explored as a potential alternative to binary logic in computer science research.

A virtue of the ternary number system is that, unlike in binary, negative numbers are naturally captured by this system without an ad hoc prefix for the sign. There is also a more economical method for rounding to a particular number of significant digits as opposed to the use of binary numbers.

Further exploration of this concept lies beyond the scope of our discussion here and will be suppressed for the sake of brevity.

3.5.1 Balanced Ternary Number System

The balanced ternary number system is a unique base-3 system that uses the digits −1, 0, and 1 to represent numbers. In this system, −1 is often denoted as "T" (for "trinary"), though alternative identifiers for −1 are in use. We will use only the designation of "T" for −1. Each digit in a balanced ternary number is referred to as a trit (analogous to how a digit in a binary number is referred to as a bit).

In the balanced ternary system, each position corresponds to a power of 3, similar to other base-n systems. The ternary point (similar to the decimal point in base-10) indicates a point where the powers of three become negative.

Example 3.49 Convert the balanced ternary number T01T to decimal form.

Solution:

$$-1 \cdot 3^3 + 0 \cdot 3^2 + 1 \cdot 3^1 + (-1) \cdot 3^0$$

$$-27 + 0 + 3 - 1$$

$$-25$$

Example 3.50 Convert the balanced ternary number 10 T1 to decimal.

Solution:

$$1 \cdot 3^3 + 0 \cdot 3^2 + (-1) \cdot 3^1 + 1 \cdot 3^0$$

$$27 + 0 - 3 + 1$$

$$25$$

Example 3.51 Convert the balanced ternary number 1 T.T1T to decimal form.

Solution:

$$1 \cdot 3^1 + (-1) \cdot 3^0 + (-1) \cdot 3^{-1} + 1 \cdot 3^{-2} + (-1) \cdot 3^{-3}$$

$$3 - 1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27}$$

$$2 - \frac{9}{27} + \frac{3}{27} - \frac{1}{27}$$

$$2 - \frac{7}{27}$$

$$1\frac{20}{27} \approx 1.\overline{740}$$

The conversion from decimal to balanced ternary is done in two steps:

a) Convert decimal to the ternary number system.
b) Convert ternary to the balanced ternary system, using the below steps:
   - traverse the ternary number, right to left, by leaving 0 and 1 as it is;
   - if we encounter 2, change it to T and add +1 to the next digit in the iteration;
   - if we encounter 3, change it to 0 and add +1 to the next digit in the iteration;
   - complete this process until you convert all the digits.

Example 3.52 Convert the decimal number 17 to balanced ternary.

Solution:

$17 \div 3 = 5$ remainder 2 (use 1)
$5 \div 3 = 1$ remainder 2 (use 1)
$1 \div 3 = 0$ remainder 1 (use 1)

Reading the digits from bottom to top, we get the ternary representation of 17 as:

17 (decimal) = 111 (ternary)

From right to left, we leave the digits as they are, since all are 1's. The balanced ternary number is the same as the regular ternary number.

Example 3.53 Convert the decimal number 238 to balanced ternary.

Solution: The ternary form of 238 is 22211 (show). Moving through the ternary number right to left, we leave the first pair of 1's as they are, then convert the 2 to T and add 1 to the digit to its left, giving us

23T11

Proceeding to the next digit, now 3, we change it to 0 and add 1 to the digit to its left, giving us

30T11

Finally, the 3 is changed to a 0, and 1 is added to the digit to its left, giving us the balanced ternary representation

100T11

Note, we can confirm this by transitioning back to decimal form:

$$1 \cdot 3^5 + 0 \cdot 3^4 + 0 \cdot 3^3 + (-1) \cdot 3^2 + 1 \cdot 3^1 + 1 \cdot 3^0$$

$$243 + 0 + 0 - 9 + 3 + 1$$

$$238$$

The balanced ternary system has some unique properties. For example, negating a number in balanced ternary is as simple as replacing each digit with its negation (-1 becomes 1, 0 remains 0, and 1 becomes -1). This property can simplify certain arithmetic operations and has been exploited in some computer designs.

Also, the balanced ternary system is a non-contiguous numeral system, meaning that not all decimal numbers have a unique finite representation. For example, the decimal number 1/2 can be represented as 0.111... (repeating) in balanced ternary.

3.6 Generalization: Base-n Number Systems

A base-n number system, also known as a radix-n system, is a positional numeral system that uses n distinct symbols to represent numbers. In a base-n system, each position corresponds to a power of n, increasing from right to left. The most common base-n systems are binary (base−2), octal (base-8), decimal (base-10), and hexadecimal (base-16), but any positive integer can be used as the base.

In a base-n system, the place values are powers of n, and the symbols used to represent the digits are typically the numbers 0 to n−1. For example, in a base-5 system, the symbols would be 0, 1, 2, 3, and 4.

The general formula for calculating the decimal value of a base-n number is:

$$\sum_{i=0}^{k-1} n \cdot b_i$$

Example 3.54 Convert the base-5 number 1234 to decimal.

Base-5 number: 1234.

Using the general formula:

$$4 \cdot 5^0 + 3 \cdot 5^1 + 2 \cdot 5^2 + 1 \cdot 5^3$$

$$4 + 15 + 50 + 125$$

$$194$$

Therefore, the decimal value of the base-5 number 1234 is 194.

To convert a decimal number to a base-n representation, we can use the division method similar to the one used for converting decimal to binary, octal, hexadecimal, or ternary. The difference (once more) is that we divide by the desired base n instead of the respective base for the aforementioned number systems.

Example 3.55 Convert the decimal number 237 to base-7.

$237 \div 7 = 33$ remainder 6
$33 \div 7 = 4$ remainder 5
$4 \div 7 = 0$ remainder 4

Reading the remainders from bottom to top, we get the base-7 representation of 237.

Base-n number systems have various applications in computer science, cryptography, and other fields. They can be used to represent numbers in a more compact form, perform arithmetic operations in different bases, or design algorithms that exploit the properties of specific number systems.

One interesting example of a base-n system is the balanced ternary system, which uses the digits −1, 0, and 1. This system has some unique properties and has been studied for its potential use in computer systems and digital signal processing.

3.7 Applications of Number Systems

Number systems play a crucial role in various fields, particularly in computer science and digital systems. In this section, we will explore some of the key applications of number systems, starting with their significance in computer architecture.

3.7.1 Number Systems in Computer Architecture

Computer architecture heavily relies on number systems to represent, store, and manipulate data. The most commonly used number systems in computer architecture are binary, octal, and hexadecimal.

3.7.1.1 Binary Number System

The binary number system is the foundation of digital computing. Computers use binary to represent and process data because electronic circuits can easily distinguish between two states: on (1) and off (0). Some key applications of binary in computer architecture include:

a) Data representation: All data in computers, including numbers, characters, and instructions, are represented using binary codes. For example, the ASCII code uses 7 bits to represent each character, while Unicode uses 16 bits or more.
b) Memory addressing: Binary numbers are used to address memory locations in a computer. Each memory cell has a unique binary address that the computer uses to access and manipulate data stored in that location.
c) Instruction encoding: Computer instructions are encoded using binary codes. Each instruction has a specific binary pattern that tells the processor what operation to perform and what data to use.

3.7.1.2 Hexadecimal Number System

The hexadecimal number system is widely used in computer architecture due to its compact representation and ease of conversion to binary. Some important applications of hexadecimal include:

a) Memory dumps: Hexadecimal is often used to display the contents of memory in a compact and readable format. Each byte (8 bits) of memory is represented by two hexadecimal digits, making it easier to analyze and debug memory-related issues.

Example 3.56 A memory dump showing hexadecimal values:
Address Hexadecimal Data

```
0x1000 7F 45 4C 46 02 01 01 00 00 00 00 00 00 00 00 00
0x1010 03 00 3E 00 01 00 00 00 90 11 00 00 00 00 00 00
```

b) Color representation: Hexadecimal is used to represent colors in various digital systems, such as HTML, CSS, and graphics software. Each color channel (red, green, and blue) is represented by a two-digit hexadecimal value, ranging from 00 to FF.

Example 3.57 The color red in hexadecimal representation:

#FF0000 (where FF represents red, 00 represents green, and 00 represents blue).

3.7.1.3 Octal Number System

While less common than binary and hexadecimal, the octal number system has some *applications* in computer architecture, particularly in Unix-based systems.

a) File permissions: Unix-based systems use octal to represent file permissions. Each permission (read, write, and execute) is represented by a three-digit octal value, with each digit representing the permissions for the owner, group, and others.

Example 3.58 File permissions in octal representation:
rwxr-xr-x is represented as 755 (where 7 = 111 in binary, representing read, write, and execute permissions for the owner; 5 = 101 in binary, representing read and execute permissions for the group and others).

3.7.2 Number Systems in Digital Logic Design

Digital logic design is the process of creating digital circuits that perform specific functions using logic gates and other components. Number systems, particularly binary, play a fundamental role in digital logic design.

3.7.2.1 Binary Number System

In digital logic design, the binary number system is used to represent the two states of a digital signal: 0 (low or false) and 1 (high or true). This representation is known as the "positive logic system." Some key applications of binary in digital logic design include:

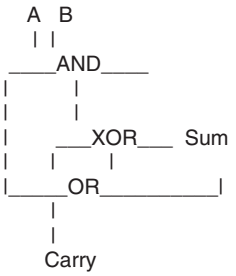
a) Logic gates: Logic gates, such as AND, OR, NOT, and XOR, operate on binary inputs to produce binary outputs. The behavior of these gates is defined using truth tables, which show the output for every possible combination of inputs.

Example 3.59 Truth table for a two-input AND gate:

Input A	Input B	Output
0	0	0
0	1	0
1	0	0
1	1	1

b) Combinational circuits: Combinational circuits are digital circuits where the output depends only on the current inputs. These circuits are built using logic gates and can perform various functions, such as arithmetic operations, data multiplexing, and code conversion.

Example 3.60 A half-adder circuit that adds two binary digits (bits) and produces a sum and a carry output:



c) Sequential circuits: Sequential circuits are digital circuits where the output depends on both the current inputs and the previous state of the circuit. These circuits use memory elements, such as flip-flops, to store the state information. The state transitions and outputs are defined using state tables and state diagrams, which represent the circuit's behavior using binary values.

Example 3.61 State table for a simple sequential circuit with one input (X) and one output (Z):

Current State	Input X	Next State	Output Z
0	0	0	0
0	1	1	0
1	0	1	1
1	1	0	1

3.7.2.2 Hexadecimal Number System

While the binary number system is the foundation of digital logic design, the hexadecimal number system is often used for convenience and readability.

a) Representation of binary data: Hexadecimal is used to represent binary data in a more compact and human-readable form. Each hexadecimal digit represents a group of four binary digits (bits), making it easier to read and write large binary numbers.

Example 3.62 A binary number represented in hexadecimal:

Binary: 1011_0110_1001_1100
Hexadecimal: B69C

b) Memory addressing: In digital systems, hexadecimal is often used to represent memory addresses. This representation makes it easier to read and manipulate memory addresses, especially when working with large address spaces.

Example 3.63 A memory address represented in hexadecimal:

0x7FFC5B7A8

Number systems, particularly binary and hexadecimal, are essential in digital logic design. They provide the foundation for representing and manipulating digital signals, designing combinational and sequential circuits, and working with digital systems at various levels of abstraction.

3.7.3 Number Systems in Cryptography

Cryptography is the practice of secure communication in the presence of adversaries. It involves techniques for encrypting and decrypting messages, as well as methods for authentication, integrity, and non-repudiation. Number systems play a crucial role in various cryptographic algorithms and protocols.

3.7.3.1 Binary Number System

The binary number system is fundamental to computer-based cryptography, as all data, including plaintext, ciphertext, and cryptographic keys, are represented and processed using binary values.

a) Symmetric encryption: Symmetric encryption algorithms, such as AES (Advanced Encryption Standard) and DES (Data Encryption Standard), operate on binary data. The plaintext is converted to binary, encrypted using a binary key, and the resulting ciphertext is also in binary format.

Example 3.64 AES encryption using a 128-bit key:

Plaintext (binary): 01101000011001010110110001101100011011111
Key (binary): 10101010101010101010101010101010101010101010101010101010101010101010
Ciphertext (binary): 11000100111010111100010011101011110001001110101111000100111010111

b) Hash functions: Cryptographic hash functions, such as SHA-256 and MD5, take binary input data of arbitrary length and produce a fixed-size binary output called a hash or digest. These functions are used for data integrity, password storage, and digital signatures.

Example 3.65 SHA-256 hash of the ASCII message "Hello, World!":

Message (ASCII): Hello, World!
Message (binary): 01001000 01100101 01101100 01101100 01101111 00101100 00100000 01010111 01101111 01110010 01101100 01100100 00100001
Hash (binary): 01011101 00000101 11100101 01100100 11100010 10011110 01000000 01100011 01000110 01010110 11001001 01110100 10011001 11010101 10010100 11110000 01111101 01010111 00100001 01001110 01110000 00001011 00111000 11001110 00111000 01000011 10110000 00010110 11111001 00100111 00000100 10001100
Hash (hexadecimal): 7f83b1657ff1fc53b92dc18148a1d65dfc2d4b1fa3d677284addd200126d9069

3.7.3.2 Modular Arithmetic and Prime Numbers

Many cryptographic algorithms, particularly those based on public-key cryptography, rely on modular arithmetic and prime numbers. These concepts are closely related to number systems and their properties.

a) RSA encryption: The RSA (Rivest–Shamir–Adleman) algorithm is a widely used public-key encryption system. It relies on the difficulty of factoring large composite numbers, which are the product of two prime numbers. The encryption and decryption processes involve modular exponentiation, where the plaintext and ciphertext are treated as integers, and the calculations are performed using modular arithmetic.

Example 3.66 RSA encryption with small prime numbers (in practice, much larger primes are used):

Public key: ($n = 33$, $e = 7$)
Private key: ($n = 33$, $d = 3$)
Plaintext: 12
Ciphertext: 12^7 mod 33 = 15
Decrypted plaintext: 15^3 mod 33 = 12

b) Diffie–Hellman key exchange: The Diffie–Hellman key exchange protocol allows two parties to establish a shared secret key over an insecure channel. The protocol relies on the difficulty of computing discrete logarithms in a finite field, which is related to modular arithmetic and prime numbers.

Example 3.67 Diffie–Hellman key exchange with small prime numbers (in practice, much larger primes are used):

Public parameters: ($p = 23$, $g = 5$)
Alice's private key: $a = 6$
Bob's private key: $b = 15$
Alice's public key: $A = 5^6 \bmod 23 = 8$
Bob's public key: $B = 5^15 \bmod 23 = 19$
Shared secret key (calculated by Alice): $S = 19^6 \bmod 23 = 2$
Shared secret key (calculated by Bob): $S = 8^15 \bmod 23 = 2$

Number systems, particularly binary, and concepts from number theory, such as modular arithmetic and prime numbers, are essential in cryptography. They provide the foundation for representing and processing data securely, as well as for designing and analyzing cryptographic algorithms and protocols.

3.8 Arithmetic in Other Number Systems

Performing arithmetic operations in number systems other than decimal can be done using similar principles and methods. In this section, we will focus on addition and subtraction in the octal number system.

3.8.1 Addition and Subtraction in Hexadecimal

The hexadecimal number system, also known as base-16, uses 16 digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, and F. The letters A through F represent the decimal values 10 through 15, respectively. When adding or subtracting hexadecimal numbers, we follow the same rules as in decimal arithmetic, but with a larger set of digits.

3.8.1.1 Addition in Hexadecimal

To add two hexadecimal numbers, we add the digits in each column, starting from the rightmost column. If the sum in any column is 16 or greater, we carry over the excess to the next column to the left.

Example 3.68 Add 7A9 (base 16) to 5C3 (base 16).

Solution:

				Carry Row
	7	A	9	
+	5	C	3	

Note: $9 + 3 = 12$, which (in base-16) corresponds to C, and we have no carry:

				Carry Row
	7	A	9	
+	5	C	3	
			C	

A and C correspond to 10 and 12, respectively, the sum of which is 22, which is 16_{16}. We write down the 6 and carry the 1:

	1			Carry Row
	7	A	9	
+	5	C	3	
		6	C	

Finally, $1 + 7 + 5 = 13$, which corresponds to D_{16}, which we write down with no carry.

	1			Carry Row
	7	A	9	
+	5	C	3	
	D	6	C	

3.8.1.2 Subtraction in Hexadecimal

To subtract two hexadecimal numbers, we subtract the digits in each column, starting from the rightmost column. If the top digit in any column is smaller than the bottom digit, we borrow from the next column to the left, keeping in mind that borrowing from the next column to the left adds 16 to our existing digit.

Example 3.69 Subtract 798_{16} from $BE5_{16}$

Solution:

	B	E	5
–	7	9	8

To subtract 8 from 5, we must borrow from the 16's place, giving us:

	B	D	21
–	7	9	8
			D

Next, we subtract 9 from D (which is 13 in decimal form), giving:

	B	D	21
–	7	9	8
		4	D

Finally, we subtract 7 from B (which is 11 in decimal form), giving a result of:

	B	D	21
–	7	9	8
	4	4	D

When performing addition and subtraction in hexadecimal, it's essential to be mindful of the base and to ensure that all digits are within the valid range (0–9 and A–F). Practice and familiarity with the hexadecimal number system will help you perform these operations more efficiently.

Example 3.70 Calculate 3B7 (base-16) + AE9 (base-16) – 5CD (base-16)

Solution:

				Carry Row
	3	B	7	
+	A	E	9	

Addition of 7 and 9 yields 16, which is 10_{16}. We write down 0 and carry the 1.

		1		Carry Row
	3	B	7	
+	A	E	9	
			0	

Addition of 1 and B and E yields $1A_{16}$, so we write down A and carry the 1.

	1	1		Carry Row
	3	B	7	
+	A	E	9	
		A	0	

Addition of 1 and 3 and A yields E, so we write down E and have no carry.

	1	1		Carry Row
	3	B	7	
+	A	E	9	
	E	A	0	

We now compute $EA0_{16}-5CD_{16}$.

	E	A	0
–	5	C	D

We must borrow from A to do our first subtraction; recalling "A" corresponds to "10" in decimal form, we are borrowing 1 of the 16 s and now have:

	E	9	16
–	5	C	D

Subtracting, we obtain:

	E	9	16
–	5	C	D
			3

To subtract C from 9, we must borrow 1 from the 256's place,

	D	25	16
–	5	C	D
		D	3

Finally, "D" corresponds to 13 in base-16, so we obtain:

	D	25	16
–	5	C	D
	8	D	3

The result is $8D3_{16}$.

3.8.2 Multiplication and Division in Hexadecimal

Multiplication and division in the hexadecimal number system follow similar principles to those in the decimal system, but with a larger set of digits (0–9 and A–F). In this subsection, we will explore these operations in detail.

3.8.2.1 Multiplication in Hexadecimal

To multiply two hexadecimal numbers, we pursue the same series of steps as was done in octal multiplication. The only distinction is the care we must take to ensure that we are using digits from the hex system.

Example 3.71 Multiplying two hexadecimal numbers $2A5_{16}$ and $1B_{16}$.

Solution:

				Carry Row for Multiplication by B
	2	A	5	
x		1	B	

B_{16} corresponds to 11 in decimal form, so $B \times 5 = 55$, which is 37_{16}. We write down the 7 and carry the 3:

		3		Carry Row for Multiplication by B
	2	A	5	
×		1	B	
			7	

B_{16} and A_{16} correspond to 11 and 10 in decimal form, the product of which is 110. Since $16 \times 6 = 96$, we see the product is $6C_{16}$, to which we add the carried 3 to give 71_{16}. We write down the 1 and carry the 7.

	7	3		Carry Row for Multiplication by B
	2	A	5	
×		1	B	
		1	7	

Next, $B_{16} \times 2_{16}$ is equivalent to $11 \times 2 = 22$ in decimal form, to which we add the carried 7. This gives us 29, which is $1D_{16}$.

	7	3		Carry Row for Multiplication by B
	2	A	5	
×		1	B	
1	D	1	7	

Turning our attention to the product of 1 with 2A5, we see we have:

	7	3		Carry Row for Multiplication by B
	2	A	5	
×		1	B	
1	D	1	7	
2	A	5		
				Sum

Finally, we add the two base-16 values together to obtain:

	7	3		Carry Row for Multiplication by B
	2	A	5	
×		1	B	
1	D	1	7	
2	A	5		
4	7	6	7	Sum

3.8.2.2 Division in Hexadecimal

As with multiplication, to divide two hexadecimal numbers, we follow a process similar to long division in the decimal system. This mimics the process used for octal numbers.

Example 3.72 Divide the hexadecimal numbers 49B7 ÷ 25.

Solution: Leaving the multiplications to you (think of the letters in their decimal form, and proceed as we demonstrated with octal division), we find:

			1	F	E
2	5	4	9	B	7
		2	5		
		2	4	B	
		−2	2	B	
			2	0	7
			−2	0	6
					1

The result is 1FE, with remainder 1.

Multiplication and division in hexadecimal may seem challenging at first, but with practice and a strong understanding of the hexadecimal number system, these operations will become more manageable. It's crucial to keep track of the base and ensure that all intermediate results and final answers are expressed in hexadecimal.

Example 3.73 Find 7C4 × D3 ÷ 1E in hexadecimal.

Solution:

		7	C	4
	×		D	3
	1	7	4	C
6	4	F	4	
6	6	6	8	C

Then,

			3	6	9	E
1	E	6	6	6	8	C
		5	A			
			C	6		
			B	4		
			1	2	8	
			1	0	E	
				1	A	C
				1	A	4
						8

The result is 369E, with remainder 8.

3.9 Floating-Point Representation

In computer systems, real numbers are typically represented using either fixed-point or floating-point formats. Floating-point representation allows for a wider range of values and higher precision compared to fixed-point representation. In this section, we will explore the differences between fixed-point and floating-point numbers and discuss how floating-point numbers are represented in binary.

3.9.1 Fixed-Point vs. Floating-Point Numbers

3.9.1.1 Fixed-Point Numbers

Fixed-point numbers are represented using a fixed number of bits for the integer part and a fixed number of bits for the fractional part. The position of the decimal point (or binary point) is fixed, hence the name "fixed-point."

Example 3.74 Fixed-point representation (8-bit, 4 bits for integer part, 4 bits for fractional part).

Integer part	Fractional part
0 1 0 1	1 0 1 0 (base 2)

In this example, the number represented is 5.625 in decimal:
Integer part: 0101 (base-2) = 5 (base-10)
Fractional part: 1010 (base-2) = 0.625 (base-10)
(0.625 = 1 × 2^(−1) + 0 × 2^(−2) + 1 × 2^(−3) + 0 × 2^(−4))

There are, as we will see in the future, advantages to fixed-point representation:

- Simple to implement and understand.
- Efficient for arithmetic operations.

Disadvantages of fixed-point representation include:

- Limited range of representable values.
- Precision is fixed and depends on the number of bits allocated for the fractional part.

3.9.1.2 Floating-Point Numbers

Floating-point numbers are represented using a sign bit, an exponent, and a mantissa (also called the significand). The decimal point (or binary point) "floats" based on the value of the exponent, allowing for a wider range of representable values and higher precision.

We will look at the particular case of IEEE 754 single-precision 32-bit representations. In this method, the representation of numbers utilizes 32 binary digits. The construction of the representation begins with a sign bit: by convention, the first bit is a sign bit (if the number is positive, the first digit is 0; if the number is negative, the first digit is 1). The remaining 31 bits are a bit more involved to find.

To find these digits, we find the binary representation of the number being converted. Once we have done this, we transform that result into the binary form of scientific notation. An example will illustrate how this is done.

Example 3.75 Find the floating-point representation (IEEE 754 single-precision, 32 bits) of 26.8.

Sign	Exponent	Mantissa
1 bit	8 bits	23 bits

Solution: The sign bit for this number will be 0, since we are starting with a positive number. The binary representation of 26 is 11010. The fractional part of the number, 0.8, is converted to binary by repeatedly multiplying by 2:

$2 \times 0.8 = 1.6$; take note of the digit to the left of the decimal point: 1.

Ignoring the whole number part of this number, we multiply the decimal part by 2:

$0.6 \times 2 = 1.2$; take note of the digit to the left of the decimal point: 1.

Repeat the process,

$0.2 \times 2 = 0.4$; take note of the digit to the left of the decimal point: 0.

$0.4 \times 2 = 0.8$; the digit to the left of the decimal point is again 0.

$0.8 \times 2 = 1.6$; the digit to the left of the decimal point is 1.

$0.6 \times 2 = 1.2$; the digit to the left of the decimal point is 1.

Notice the digits will begin repeating at this point. The digits to the right of the binary point will be 1100110011001100…

Combining the whole and fractional parts of our results, we have the binary representation of 26.8 as 11010.1100110011001100…

We will now convert this value to binary scientific notation. The binary point is moved to the left until a single digit lies to its left and multiplied by the appropriate power of 2:

$1.10101100110011001100 \times 2^4$

The number of places the binary point was moved is four. Since it is moved to the left, the exponent on 2 is positive.

Recall the formatting requirement for our representation:

Sign	Exponent	Mantissa
1 bit	8 bits	23 bits

For us, the sign bit was 0. To find the 8 bits corresponding to the exponent, we add the power on 2 to the bias, which was 127. So, $127 + 4 = 131$, which (in binary) is 10000011.

The remaining 23 bits are the first 23 digits following the binary point in the binary scientific notation form we found, $1.10101100110011001100 \times 2^4$. Since it is a repeating decimal, we write the first 23 digits of the pattern:1010 1100 1100 1100 1100 110

Putting all of this together, we find the floating-point representation (IEEE 754 single-precision, 32 bits) of 26.8 is:

0100 0001 1101 0110 0110 0110 0110 0110

The value of a floating-point number is calculated as follows:

(-1)^sign $\times (1 +$ mantissa$) \times 2$^(exponent−bias)

where the bias is 127 for single-precision floating-point numbers.

Example 3.76 Find the floating-point representation of −0.15625.

Solution: Since the sign of the number is negative, the sign bit's value will be 1. The whole number part of our given value is 0, so it remains to find the fractional part of the number, in binary.

$0.15625 \times 2 = 0.3125$; we note the digit to the left of the decimal point, 0.

$0.3125 \times 2 = 0.625$; we again note the digit to the left of the decimal point, 0.

$0.625 \times 2 = 1.25$; we note the value of 1.

$0.25 \times 2 = 0.5$; we note the value of 0.

$0.5 \times 2 = 1.0$; we note the value of 1, and the binary fractional representation terminates at 0.00101_2.

Combining the whole and fractional parts, our binary representation of 0.15625 is 0.00101_2. We convert this to binary scientific notation and obtain 1.01×2^{-3}. We add the exponent of 2 to the bias, obtaining $127 + (-3) = 124$, which (in binary) is 1111100.

Putting everything together, we begin with the sign bit, which is 1, followed by the exponent part of the number (expressed as 8 digits) 01111100, and followed by 23 digits from the fractional part of the scientific notation (padded out with zeros, as needed), 01000000000000000000. Our floating-point representation of 0.15625 is therefore:

1011 1110 0010 0000 0000 0000 0000 0000

Among the advantages of floating-point representation are:

- Wide range of representable values.
- High precision for small and large numbers.

Disadvantages of floating-point representation:

- More complex to implement and understand compared to fixed-point.
- Rounding errors can occur due to the finite precision of the mantissa.

Floating-point representation is the most common choice for representing real numbers in modern computer systems. The IEEE 754 standard defines the format for single-precision (32 bits) and double-precision (64 bits) floating-point numbers, which are widely used in scientific computing, financial calculations, and other applications requiring high precision.

3.9.2 IEEE 754 Floating-Point Standard

The IEEE 754 standard is the most widely used floating-point representation in modern computer systems. It defines the formats for single-precision (32 bits) and double-precision (64 bits) floating-point numbers, as well as the rules for arithmetic operations and rounding.

Single-Precision Floating-Point Format (32 bits):

Sign	Exponent	Mantissa
1 bit	8 bits	23 bits

Double-Precision Floating-Point Format (64 bits):

Sign	Exponent	Mantissa
1 bit	11 bits	52 bits

The value of a floating-point number is calculated as follows:

$$(-1)\text{^sign} \times (1 + \text{mantissa}) \times 2\text{^(exponent} - \text{bias)}$$

where the bias is 127 for single-precision and 1023 for double-precision floating-point numbers.

Example 3.77 Single-precision floating-point representation of 0.1.

Solution:

Sign: 0 (positive)
Exponent: 01111011 (base-2) = 123 (base-10)
Mantissa: 10011001100110011001101 (base-2)

Calculating the value:

$$\begin{aligned}&(-1)^0 \times (1 + 0.600000023841857910156 25) \times 2^{(123-127)}\\ &\quad = 1 \times 1.600000023841857910156 25 \times 2^{(-4)} \approx 0.1\end{aligned}$$

Note that the binary representation of 0.1 is not exact, as it has an infinite binary expansion. The IEEE 754 standard rounds the value to the nearest representable number.

Example 3.78 Double-precision floating-point representation of −3.14159265358979323846.

Solution: Sign: 1 (negative)

Exponent: 10000000000 (base-2) = 1024 (base-10)
Mantissa: 1001001000011111101101010100010001000010110100011000 (base-2)

Calculating the value:

$$\begin{aligned}&(-1)^1 \times (1 + 0.570288748756655454153914776947383293310856816 8) \times 2^{(1024-1023)}\\ &\quad \approx -3.14159265358979323846\end{aligned}$$

3.9.2.1 IEEE 754 Floating-Point Arithmetic

The IEEE 754 standard also defines the rules for arithmetic operations on floating-point numbers, such as addition, subtraction, multiplication, and division. These rules ensure that the results of arithmetic operations are consistent across different implementations and platforms.

Example 3.79 Addition of two single-precision floating-point numbers

Number 1:010000010 01000000000000000000000 (base-2) = 3.125
Number 2: 010000001 10100000000000000000000 (base-2) = 2.625

To add these numbers, we first align the binary points by adjusting the exponents:
010000010 01000000000000000000000
010000001 10100000000000000000000

010000010 01010000000000000000000
The result is 010000010 01010000000000000000000 (base-2), which represents the decimal value 5.75.

3.9.2.2 Rounding Modes

The IEEE 754 standard defines four rounding modes for arithmetic operations:

- Round to nearest (default): Round the result to the nearest representable number. If the result is exactly halfway between two representable numbers, round to the one with the least significant bit equal to zero.
- Round toward $+\infty$: Round the result to the nearest representable number that is greater than or equal to the exact result.
- Round toward $-\infty$: Round the result to the nearest representable number that is less than or equal to the exact result.
- Round toward zero: Round the result to the nearest representable number that is closer to zero than the exact result.

Example 3.80 Rounding single-precision floating-point numbers
Solution:

Exact value: 1.1 (base-10) = 1.00011001100110011001100110011001100110011001100110011... (base-2)

Rounded values (single-precision):

Round to nearest: 001111111 00011001100110011001101 (base-2) $\approx$ 1.100000023841858
Round toward $+\infty$: 001111111 00011001100110011001101 (base-2) $\approx$ 1.100000023841858
Round toward $-\infty$: 001111111 00011001100110011001100 (base-2) $\approx$ 1.099999904632568
Round toward zero: 001111111 00011001100110011001100 (base-2) $\approx$ 1.099999904632568

The IEEE 754 floating-point standard has been widely adopted in modern computer systems and provides a consistent and efficient way to represent and manipulate real numbers.

3.9.2.3 Single-Precision and Double-Precision Formats

The IEEE 754 standard defines two main formats for floating-point numbers: single-precision (32 bits) and double-precision (64 bits). These formats differ in the number of bits allocated for the sign, exponent, and mantissa, which affects the range and precision of the representable values.

Single-Precision Format (32 bits) The single-precision format uses 32 bits, divided into three parts:

Sign	Exponent	Mantissa
1 bit	8 bits	23 bits

Sign bit: 0 for positive numbers, 1 for negative numbers
Exponent: 8 bits, biased by 127 (actual exponent = stored exponent – 127).
Mantissa: 23 bits, representing the fractional part of the significand (implied leading 1).
The value of a single-precision floating-point number is calculated as follows:

$(-1)\text{^sign} \times (1 + \text{mantissa}) \times 2\text{^(exponent} - 127)$

Example 3.81 Find the single-precision representation of 42.625

Solution:

Sign: 0 (positive)
Exponent: 10000100 (base 2) = 132 (base 10)
Mantissa: 01010100000000000000000 (base 2)

Calculating the value:

$$(-1)^0 \times (1 + 0.32) \times 2^{(132-127)} = 1 \times 1.32 \times 2^5 = 42.625$$

Double-Precision Format (64 bits) The double-precision format uses 64 bits, divided into three parts:

Sign	Exponent	Mantissa
1 bit	11 bits	52 bits

Sign bit: 0 for positive numbers, 1 for negative numbers
Exponent: 11 bits, biased by 1023 (actual exponent = stored exponent − 1023)
Mantissa: 52 bits, representing the fractional part of the significand (implied leading 1)
The value of a double-precision floating-point number is calculated as follows:

(−1)^sign × (1 + mantissa) × 2^(exponent − 1023)

Example 3.82 Find the double-precision representation of −3.14159265358979323846

Solution:

Sign: 1 (negative)
Exponent: 10000000000 (base 2) = 1024 (base 10)
Mantissa: 1001001000011111101101010100010001000010110100011000 (base 2)

Calculating the value:

$$(-1)^1 \times = (1 + 0.5702887487566554541539147769473832933108568168) \times 2^{(1024-1023)}$$
$$\approx -3.14159265358979323846$$

Comparison of Single-Precision and Double-Precision Single-precision and double-precision formats have different characteristics in terms of range and precision:

Single-Precision:

- Exponent range: −126 to +127.
- Significand precision: 24 bits (23 bits stored, 1 implied).
- Range: ±1.18 × 10^−38 to ±3.40 × 10^38.
- Precision: about seven decimal digits.

Double-Precision:

- Exponent range: −1022 to +1023.
- Significand precision: 53 bits (52 bits stored, 1 implied).
- Range: ±2.23 × 10^−308 to ±1.80 × 10^308.
- Precision: about 15 decimal digits.

Example 3.83 Comparing single-precision and double-precision representations of 0.1 (base 10).

Solution: Single-precision representation:

$$001111011\ 10011001100110011001101 \text{ (base 2)} \approx 0.100000001490116119384765625$$

Double-precision representation:

$$001111111011\ 1001100110011001100110011001100110011001100110011010 \text{ (base 2)}$$
$$\approx 0.1000000000000000005551115123125782702118158340454101562 5$$

As seen in the example, the double-precision representation of 0.1 is closer to the actual value than the single-precision representation, due to the higher precision provided by the additional bits in the mantissa.

When choosing between single-precision and double-precision formats, consider the following factors:

- Required range and precision for the application.
- Memory constraints (double-precision numbers require twice as much storage as single-precision numbers):
  - Performance considerations (single-precision arithmetic may be faster on some hardware).

In most cases, double-precision is the default choice for scientific and engineering applications, while single-precision may be sufficient for certain graphics and gaming applications where memory bandwidth and performance are critical.

4

Straight-Line Equations and Graphs

4.1 Introduction to Straight Lines

In mathematics, the Cartesian coordinate system provides a fundamental framework for graphing lines and visualizing relationships between two variables. Named after the mathematician René Descartes (1596–1650), this system uses two perpendicular axis lines (together referred to as axes [pronounced: ak sees]), the x-axis (horizontal) and y-axis (vertical), intersecting at a point referred to as the origin of the plane, identified as (0, 0).

The axes are number lines, with the positive values of the x-axis extending to the right, negative values to the left, and the positive values of the y-axis extending upward, negative values downward. The plane itself is a two-dimensional surface (it has no thickness whatsoever) with no boundary.

It is useful to visualize the plane as a flat surface, like a tabletop or a sheet of paper, but this serves only as a reference for us. The plane, remember, extends forever, while the tabletop or paper has only a finite surface area. Likewise, the plane has no thickness, while the tabletop and the sheet of paper do have thickness.

We present the content of this chapter with the presumption that terminology and fundamental notions of the Cartesian plane (the aforementioned x-axis, y-axis, origin, identification of the location of points using ordered pair coordinates, the quadrants, and so on) are well-understood. If you should need a review of these concepts, a text in college algebra presents these topics as primitives and develops the structure in detail.

When looking at the plane, your eye should always travel from ***left to right***. This is a convention universally accepted, and so we consider objects in the plane beginning at the left of the plane and moving to the right. The utility of doing so allows us to develop an intuitive understanding of concepts like "increasing" or "decreasing" quantities.

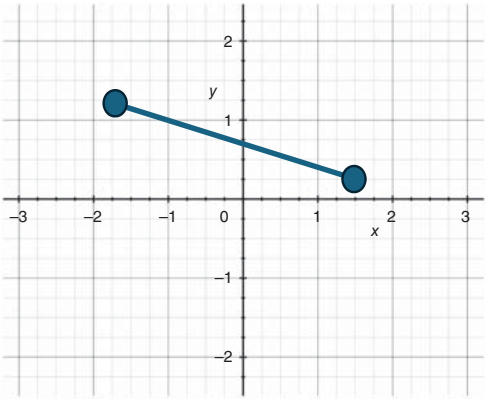
4.2 Definition of a Straight Line

A straight line is the simplest type of geometric shape, defined as the set of all points that lie along a continuous, non-curved path passing through two distinct points in the plane. They, as was the case with the plane, are considered mathematical "abstractions," because we cannot truly draw a mathematical straight line, as it has no thickness.

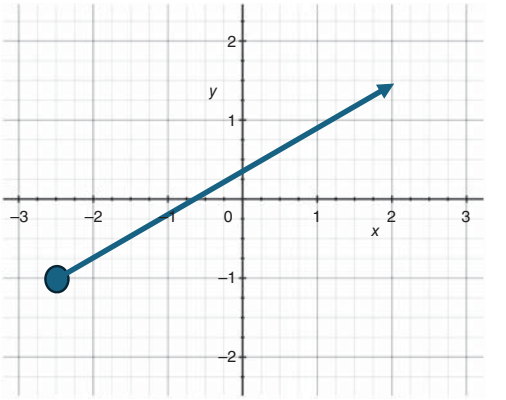
A line has no endpoints; that is, it continues forever along a non-curved path in the plane. If a non-curved path has finite length, extending from a particular point in the plane to another, we refer to it as a line segment, and if the non-curved path extends forever in one direction from a particular

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

starting point in the plane, we refer to it as a half-line. The points at which a line segment begins and ends are called endpoints of the line segment, and the starting point of a half-line is called its initial point.



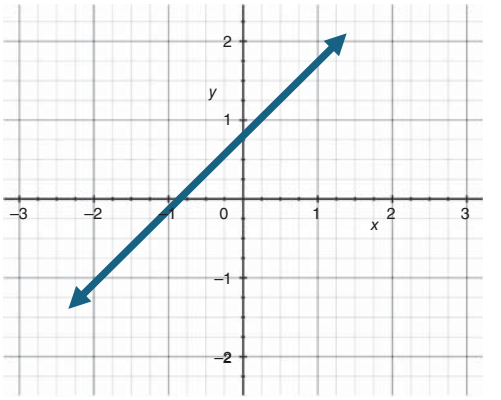
A line segment



A half-line

The circles at the end of the segment identify the endpoints, while in the half-line, it indicates the initial point. The arrowhead on the half-line indicates continuation forever, following the same direction as shown.

A straight line, in contrast to these two illustrations, continues forever in BOTH directions.



A Line

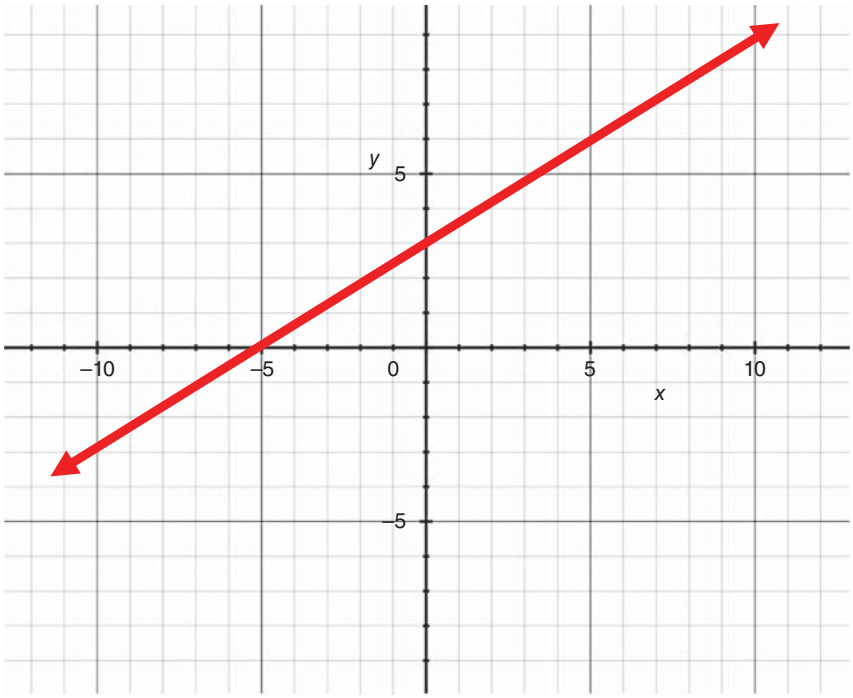
Hereafter, when we use the term "line," we mean a straight line that extends without limit along a non-curved path. Care should be taken when specifically referring to a line segment or a half-line. It is common, for instance, for individuals to refer to a line segment as a "line," but this is incorrect.

If A and B are points on a line, we typically designate the line $\overleftrightarrow{AB}$. A line segment having endpoints A and B is identified as $\overline{AB}$, and a half-line with initial point A and additional point B on it is represented as $\vec{AB}$. When speaking of these more generally, we often simply refer to it as "the line/line segment/half-line L."

4.3 The Intercepts of a Line

Lines may or may not cross the axis lines of the plane at individual points. If they do, we identify the points where this occurs as the x-intercept and the y-intercept, where the letter identifies the axis being crossed. It is possible the two points would coincide; this is the situation where the line happens to pass through the origin of the plane. Keep in mind the intercept is a single point, which leads to a "special case" we will address shortly.

The intercepts are referenced as significant graphing points in many applications of linear equation graphs, and so we shall specifically state the y-intercept of a line will be identified by the ordered pair $(0, b)$. We can refer to the x-intercept as $(a, 0)$, although the nomenclature here is not standardized.



A line having x-intercept $(-5, 0)$ and y-intercept $(0, 3)$.

It is possible a line could "coincide" with one of the coordinate axes. That is, the line could lie directly upon the x-axis or the y-axis. In such a case, we speak of only one type of intercept, which occurs at the origin.

For a vertical line coincident with the y-axis, the x-intercept is the origin; the y-axis is met at every point on the y-axis, and so there is no distinct y-intercept.

For a horizontal line coincident with the x-axis, the y-intercept is the origin, and there is no distinct x-intercept.

4.4 Slope of a Line

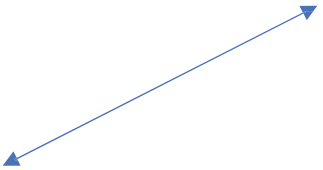
A defining characteristic of a line is its slope, which we represent by the letter ***m***. The slope provides a numerical description of the steepness of a line or the extent to which a line is tilted from horizontal, as well as the direction of such tilt. This description, by implication, tells us a horizontal line in the plane must have slope $m = 0$, since it has no tilt from horizontal.

If two distinct lines are under consideration, we typically identify them using subscripts, as L_1 and L_2. In such a case, the slopes of the line are also identified using the subscript associated with the line it describes, m_1 and m_2.

The greater the absolute value of the slope of a line, the steeper the line will be. The extreme case of steepness, a vertical line, will be considered shortly.

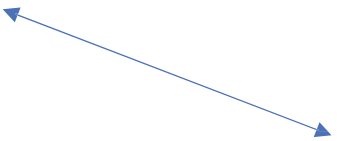
Frequently, it is merely the sign of m that is of importance in a particular problem or scenario. Keeping in mind the practice we mentioned earlier of viewing the plane from left to right, we may introduce the notion of an increasing line and a decreasing line.

Increasing lines tilt upward as you look at them from left to right:



Such lines have a positive slope.

Decreasing lines tilt downward as you look at them from left to right:



Such lines have a negative slope.

The slope, m, of a straight line is sometimes referred to using the visual references of "rise over run" and is calculated as the change in the y-coordinate divided by the change in the x-coordinate between any two points on the line. Mathematically, the slope can be expressed as:

$$m = (y_2 - y_1)/(x_2 - x_1)$$

where (x_1, y_1) and (x_2, y_2) are any two distinct points on the line.

Example 4.1 Suppose a line passes through $(-3, 8)$ and $(5, 1)$. What is the slope of this line?

Solution: While it does not matter (for the purpose of computation) which point is (x_1, y_1), we shall typically associate the first point given with this designation. The purpose of our doing so is to emphasize the notion that there is value in having a systematic approach to working on problems. If we develop consistency in our approach, we can have "one less thing to worry about" in our work.

In this first example, we will demonstrate the same result for the slope is obtained no matter which point is identified as (x_1, y_1).

We first identify $(x_1, y_1) = (-3, 8)$, and $(x_2, y_2) = (5, 1)$. With these designations, we use the slope formula to obtain:

$$\mathrm{m} = \frac{1-8}{5-(-3)} = \frac{-7}{8} = -\frac{7}{8}$$

Alternatively, had we said $(x_1, y_1) = (5, 1)$ and $(x_2, y_2) = (-3, 8)$, we would have found:

$$\mathrm{m} = \frac{8-1}{-3-5} = \frac{7}{-8} = -\frac{7}{8}$$

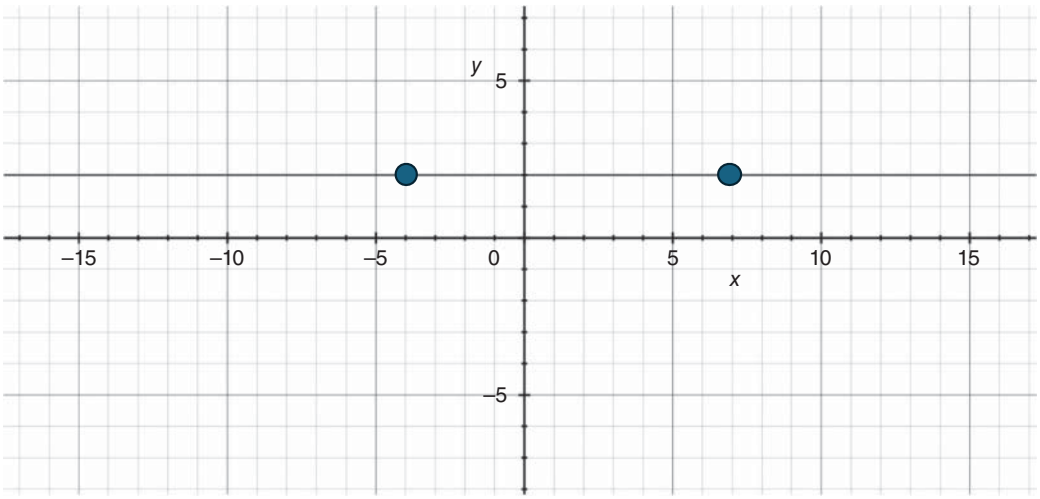
As mentioned earlier, both calculations yield the same value for m, and so the choice truly is arbitrary.

Example 4.2 Suppose a line passes through the points (7,2) and (−4,2). What is its slope?

Solution: We first identify $(x_1, y_1) = (7, 2)$, and $(x_2, y_2) = (-4, 2)$. With these designations, we use the slope formula to obtain:

$$\mathrm{m} = \frac{2-2}{-4-7} = \frac{0}{-11} = 0$$

This result catches the eye, because anytime zero appears as a result, we should wonder why it is occurring. Graphing this line (by plotting the points and producing a sketch), we find:



We observe the line is horizontal; that is, there is no vertical change in the line as our eye follows it from left to right. Recall we had mentioned earlier a horizontal line has slope $m = 0$, and this claim is borne out by the example.

Example 4.3 Suppose a line passes through $\left(2, \frac{4}{5}\right)$ and (−4, 6). What is the slope of this line?

Solution: The first point given is $\left(2, \frac{4}{5}\right)$, so we identify (x_1, y_1) as $\left(2, \frac{4}{5}\right)$ and (x_2, y_2) as (−4, 6). Substitution into the formula for slope yields the complex fraction:

$$\mathrm{m} = \frac{6 - \frac{4}{5}}{-4-(2)}$$

There are multiple methods for simplifying such a fraction, but perhaps the most straightforward is to find the least common denominator of all the fractions within the overall expression (here, i.e., 5) and to multiply all terms by that value. This will transform the complex fraction into an ordinary fraction, more easily simplified.

$$\mathrm{m} = \frac{5 \cdot 6 - 5 \cdot \frac{4}{5}}{5 \cdot (-4) - (5 \cdot 2)} = \frac{30-4}{-20-10} = \frac{26}{-30} = -\frac{13}{15}$$

Example 4.4 Suppose a line passes through $(-5, 3)$ and $(4, 7)$. What is the slope of this line?

Solution: Following our convention for identifying points, we have (x_1, y_1) as $(-5, 3)$ and (x_2, y_2) as $(4, 7)$. Substitution into the formula for slope yields the fraction:

$$m = \frac{7-3}{4-(-5)} = \frac{4}{9}$$

Example 4.5 Suppose a line passes through $(4, 3)$ and $(4, -5)$. What is the slope of this line?

Solution: Again, we assign the first point the label (x_1, y_1) and find:

$$m = \frac{3-(-5)}{4-(4)} = \frac{8}{0}$$

Division by zero, recall, is a concept that is meaningless. In mathematics, we refer to such concepts as being "undefined." The slope of a vertical line is always undefined, because the denominator in the formula for slope will always simplify to 0.

It is relevant at this point to emphasize precision in our choice of words. It would be ambiguous to refer to either a horizontal or a vertical line as having "no slope." This phrase places us on a fence where our meaning is unclear. Does "no slope" refer to "zero slope," or does it refer to "undefined slope?" Clearly, this vagueness is unacceptable, and so we insist upon the specific use of the terms "zero slope" and "undefined slope."

4.4.1 The Rise over Run Interpretation of the Slope of a Line

Earlier, we mentioned that slope is referred to using the visual interpretation "rise over run." Before proceeding, let us review what is meant by this statement. If the slope value is viewed as a fraction, perhaps given in that form or recognized as an equivalent form of an integer (recall, for instance, the integer 7 is equivalent to $\frac{7}{1}$), the fraction's numerator is referred to as the "rise," its denominator as the "run." As we move from left to right in the plane from one point on the line to another, the rise is the vertical change in our location (accepting that a positive rise is considered upward movement, while a negative rise is considered downward), while the run is the horizontal change to the right.

For example, suppose a line is known to have slope $m = \frac{2}{5}$; the rise is 2 units upward, the run is 5 units to the right. If the slope were known to be $m = -\frac{9}{13}$ (adopting the convention, the negative sign will be associated with the numerator of the fraction), the rise is -9 units downward, and the run is 13 units to the right.

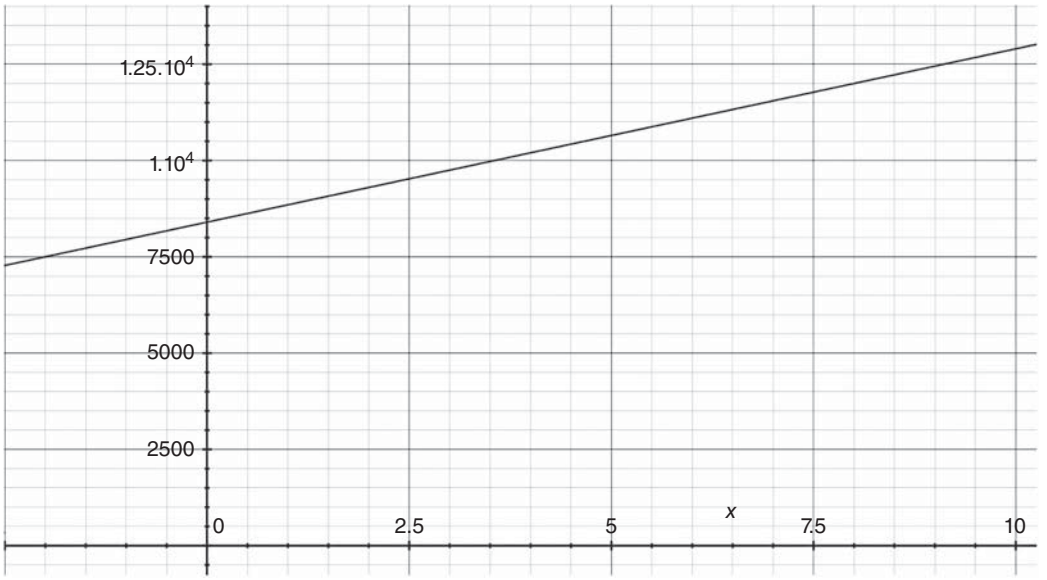
We note the run is, in some works, allowed to be positive or negative. We believe this introduces an unnecessary level of complexity, and so we will insist the run will always be to the right.

4.4.2 Slope as a Rate of Change

When we encounter lines in applications, there is a common interpretation of slope as the "rate of change" in a value. We are familiar with this concept from everyday life, as (when driving) we understand our rate of travel is ___ miles per hour. The word "per" indicates this is a rate of change.

The rate of change is always given as ___ (units of the vertical axis) per (units of the horizontal axis). That is, if we have a scenario described (for some reason) by a line, we can use the slope to produce the rate of change of the vertical axis quantity.

Imagine we have a graph for the in-state tuition at a particular public 4-year college:



The vertical axis represents the tuition (in dollars), the horizontal axis the number of credits beyond 12 taken by a student. The slope (viewed as a rate of change) would have units of dollars per credit hour past 12.

4.5 The Equation of a Line

Straight lines can be described mathematically using the equation $y = mx + b$, where m represents the slope (or rate of change) of the line and b represents the y-intercept (the point where the line intersects the y-axis, having coordinates $(0, b)$). This equation can be reorganized into various other forms each of which is equivalent to $y = mx + b$ and which will be discussed shortly.

A common description of such equations is that "x and y appear to the power 1, only," in linear equations. This can be somewhat misleading, as an example such as $y = \frac{5}{x+1}$ has both x and y to the power 1, yet its graph is not a straight line in the plane. We will specifically demand a linear equation is one that can be arranged in the format $y = mx + b$.

Despite the fact the equation can be algebraically reorganized, there is a unique equation that represents any given line. The $y = mx + $ b form is generally the most useful and familiar of the equation forms, and it is given the specific name "slope–intercept form." The choice of name is motivated by the ability (when given this form) to identify the slope and the y-intercept by inspection. If the sign between mx and b should be subtraction, recall the equation can be viewed as $y = mx + (-b)$.

Example 4.6 The following equations of lines are in the slope–intercept form; identify the slope, m, and the y-intercept $(0, b)$.

a. $y = 3x + 8$
b. $y = 4x - 10$
c. $y = \frac{1}{5}x - 2$

Solution:

a) The slope is $m = 2$, the y-intercept is $(0, 8)$.
b) The slope is $m = 4$, the y-intercept is $(0, -10)$.
c) The slope is $m = \frac{1}{5}$, the y-intercept is $(0, -2)$.

When we state, "the equation of a line can be reorganized," we mean we can manipulate the equation using any of the operations (addition, subtraction, multiplication, and division) of basic algebra, using the Principles of Equality.

For instance, consider the line given in the slope–intercept form as $y = 4x - 12$. Although it seems contrived to do so, we could use the additive identity property of algebra to rewrite the left-hand side of the equation as $y - 0$ to obtain $y - 0 = 4x - 12$. Further, we could use the distributive law to "factor" the right-hand side of the equation to obtain $y - 0 = 4(x - 3)$.

This rewrite of the equation unexpectedly takes us to a notable form of the equation of a line, called the "**point–slope form**" of the line's equation. This form is useful for finding the equation of a line from its graph (or the coordinates of a single point on its graph).

Of course, if the y-intercept is clearly visible on the graph and the slope easily determined, there would be no need for the point–slope form. It is for the cases where this is not possible that we introduce this format.

There are two possible variations of this problem:

- We can be given TWO points on the line, which we can use to find the slope m; after which we can disregard one of the two given points and use the slope you found and the remaining single point to find the equation.
- You could be given the graph of the line, from which you could identify two points on the line, and then you could find the slope, then disregard one of the two points and use the remaining single point and the slope to find the equation.

The point–slope formula tells us if a line's slope is m and it passes through a point (x_1, y_1), then:

$$\boldsymbol{y} - y_1 = (\boldsymbol{x} - x_1)$$

Note the similarity of this form to the reorganization we produced not long ago:

$$\boldsymbol{y} - 0 = 4(\boldsymbol{x} - 3)$$

The variables "y" and the "x" are in bold because we wish to emphasize these will remain variables throughout this process—we are going to insert values for m, y_1, and x_1. In the example above, the value of m corresponds to the location of the number 4, and a point through which the line passes has coordinates $(0, 3)$.

Example 4.7 Find the equation of the line passing through $(-3, 8)$ having slope $m = \frac{2}{9}$.

Observe here that $(x_1, y_1) = (-3,8)$, and that m is given. Substitution into the point–slope formula yields:

$$y - 8 = \frac{2}{9}(x - (-3))$$

While this is an acceptable form of the equation of the line, let us practice some algebra by reorganizing the equation to the slope–intercept form.

$$y - 8 = \frac{2}{9}(x + 3)$$

$$y - 8 = \frac{2}{9}x + \frac{2}{3}$$

$$y = \frac{2}{9}x + \frac{2}{3} + 8$$

$$y = \frac{2}{9}x + \frac{26}{3}$$

From here, we see the slope is $\frac{2}{9}$ and the y-intercept is $\left(0, \frac{26}{3}\right)$. Parenthetically, observe the line is tilted upward as we go from left to right in the plane.

Example 4.8 Find the equation of the line passing through (2, 4) and (5, 8).

Since we wish to find the equation of the line, the natural choice is to use the point–slope formula. However, this requires us to know a point on the line (we actually have two points here!) and the slope of the line. When two points are given, we can *find* the slope, then use it and one of the two given points to find the equation of the line.

Using the slope formula:

$$\mathrm{m} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{8 - 4}{5 - 2} = \frac{2}{3}$$

Substitute the slope and one of the given points (e.g., (2, 4)) into the point–slope form of the equation of a line:

$$y - 4 = \frac{4}{3}(x - 2)$$

As in the previous example, this could be simplified to the slope–intercept form algebraically:

$$y - 4 = \frac{4}{3}x - \frac{8}{3}$$

$$y = \frac{4}{3}x - \frac{8}{3} + 4$$

$$y = \frac{4}{3}x + \frac{4}{3}$$

We notice how a potential challenge now arises for us: in each of the examples just shown, we have two equations (apparently) for the same line, yet we know the equation of a line is unique. If we were given the two equations from the first example:

$$y - 8 = \frac{2}{9}(x - (-3))$$

and

$$y = \frac{2}{9}x + \frac{26}{3}$$

how would we be able to tell they were both describing the same line? To resolve this, we must agree upon a form of the equation of a line, which we shall identify as the "Standard Form." This is a form of the equation we shall agree to use for the purpose of comparison. If there is a question whether two equations represent the same line, that question will be resolved through a conversion to the standard form.

4.6 Standard Form of a Straight-Line Equation

The standard form of a straight-line equation is given by the expression:

$$Ax + By = C$$

where A, B, and C are real numbers, and A and B are not both 0, A is positive, and A, B, and C have no common factors. This form is useful for graphing lines and determining certain properties of the line, such as the x- and y-intercepts.

The restrictions imposed upon A, B, and C ensure the uniqueness of the standard form.

Let us again consider the equations:

$$y - 8 = \frac{2}{9}(x - (-3))$$

and

$$y = \frac{2}{9}x + \frac{26}{3}$$

We will show that these represent the same line by converting each to the standard form and comparing the results.

$$y - 8 = \frac{2}{9}(x - (-3))$$

$$9(y - 8) = 9 \cdot \frac{2}{9}(x - (-3))$$

$$9y - 72 = 2x + 6$$

$$-2x + 9y = 78$$

$$2x - 9y = -78$$

Next,

$$y = \frac{2}{9}x + \frac{26}{3}$$

$$9y = 9(2/9\,x + 26/3)$$

$$9y = 2x + 78$$

$$-2x + 9y = 78$$

$$2x - 9y = -78$$

Both equations simplify to the same standard form and thus represent the same line. It is important to realize that the true utility of the standard form is this ability for comparison of forms and to ensure the equations "match up."

4.7 The Primary Uses of the Three Forms of the Equation of a Line

The existence of alternative forms of the equation of a line can lead to confusion as to which form should be used. Unless told specifically to obtain a particular form, the primary motivation for the choice to use a particular form of a line's equation can be synopsized:

1) Slope–intercept form is useful if one wishes to know the slope and the y-intercept of the line. It is also very convenient to produce the graph of the line from this form of its equation.
2) Point–slope form is used when information about the line is in hand (two points on it, a point on it and the slope of the line, or the graph), and the equation of the line is being sought.
3) Standard form is useful as a point of reference at which different reorganizations of the equation of the line must be compared to determine if the equations all represent the same line. It is also relatively easy to graph from this form.

4.8 Graphing Straight Lines

Straight lines can be graphed on the Cartesian coordinate system using various methods, one of which is "point-plotting." This process involves the creation of a "table of values," which allows us to identify several points on the graph. We can use the equation $y = mx + b$ to create the table of values, from which we obtain the coordinates of and plot points on the coordinate plane, then sketch the line passing through those points.

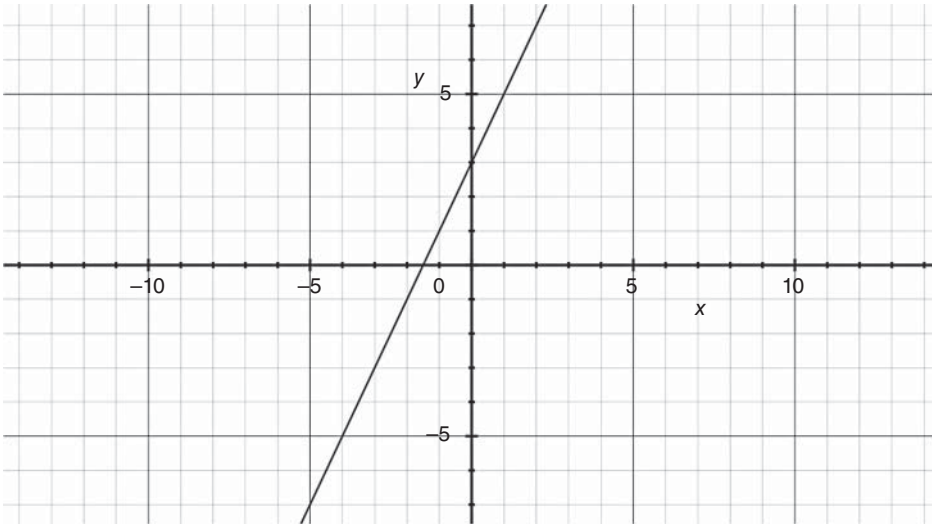
You may have heard the expression: "Two points determine a line." The meaning of this is not clear to all students, so we will elaborate upon that just a bit. If we were to plot a single point on the Cartesian plane and then be asked to draw the line through that point, we would face a significant challenge— "the" is a word that implies uniqueness. In other words, we are asking for the ONLY line that passes through that point. However, given a single point in the Cartesian plane, there are infinitely many possible lines passing through that point, and so we cannot guarantee uniqueness. On the other hand, if we were given two distinct points, we would not face this challenge. There is a unique line that passes through the two distinct points.

In the process of graphing, it is customary to find three points (with the third point serving as a "check" of our work). When we find three points and plot them, they must align on a straight-line path in the plane. If they do not, we have made a computational error and must check our work. Obviously, the challenge there is that one or more of the points could have been plotted incorrectly, and so we must check them all, to be certain.

For example, consider the graph of the equation $y = 2x + 3$. We can plot several points on this graph by plugging in different values of x and solving for the corresponding values of y:

- When $x = 0$, $y = 2(0) + 3 = 3$, so the point $(0, 3)$ lies on the graph.
- When $x = 1$, $y = 2(1) + 3 = 5$, so the point $(1, 5)$ lies on the graph.
- When $x = -1$, $y = 2(-1) + 3 = 1$, so the point $(-1, 1)$ lies on the graph.

By plotting several such points, we can see the points fall on a straight-line path on the coordinate plane. The resulting graph is shown below.



In the above figure, we have plotted the line $y = 2x + 3$ on the coordinate plane. The slope of the line is 2, and the y-intercept is 3.

The values for x chosen in the example were $x = 0$, 1, and -1. These are extremely good choices for x in the case where m is an integer.

Vertical and horizontal lines are special cases we consider in isolation:

Vertical lines have equation form $x = \mathrm{K}$, where K is some number. The reason for this becomes apparent if one graphs a vertical line and picks out distinct points lying on it. It is immediately observed all the points on the line possess the same x-coordinate, and this is precisely the value K. Possibly, you would have to re-write the equation to put it in that form. The standard form of the equation of such a line is $x + 0y = \mathrm{K}$.

The vertical line through the point (5,3), for instance, has equation $x = 5$, or (in the standard form) $x + 0y = 5$.

Horizontal lines have equation form $y = \mathrm{K}$, where K is some number. As mentioned in the case of vertical lines, if a horizontal line is graphed in the plane and distinct points on it are identified, it becomes clear the points all have the same y-coordinate, the value K. Possibly, you would have to rewrite the equation to put it into that form. The standard form of the equation of such a line is $0\mathrm{x} + \mathrm{y} = \mathrm{K}$.

The horizontal line through the point (5, 3) has equation $y = 3$, or (in the standard form) $0\mathrm{x} + \mathrm{y} = 3$.

4.9 Graphing Using the Standard Form

The standard form of a straight-line equation is given by:

$$\mathrm{A}x + \mathrm{B}y = \mathrm{C}$$

where A, B, and C are real numbers, A is non-negative, A and B are not both 0, and A, B, and C have no common factors other than 1. While the slope–intercept form is more convenient for graphing, we can still graph lines in the standard form by following these steps:

- Identify the x-intercept and y-intercept of the line.
- Plot the x-intercept and y-intercept points on the coordinate plane.
- Draw the line passing through the two intercept points.

To find the x-intercept and y-intercept, we can use the following formulas:

x-intercept: $(C/A, 0)$.
y-intercept: $(0, C/B)$.

Example 4.9 Graph the line with equation $2x - 3y = 12$.

Solution: To find the x-intercept, we set $y = 0$ and solve for x:

$2x - 3(0) = 12$
$2x = 12$
$x = 6$

The x-intercept is $(6, 0)$.

To find the y-intercept, we set $x = 0$ and solve for y:

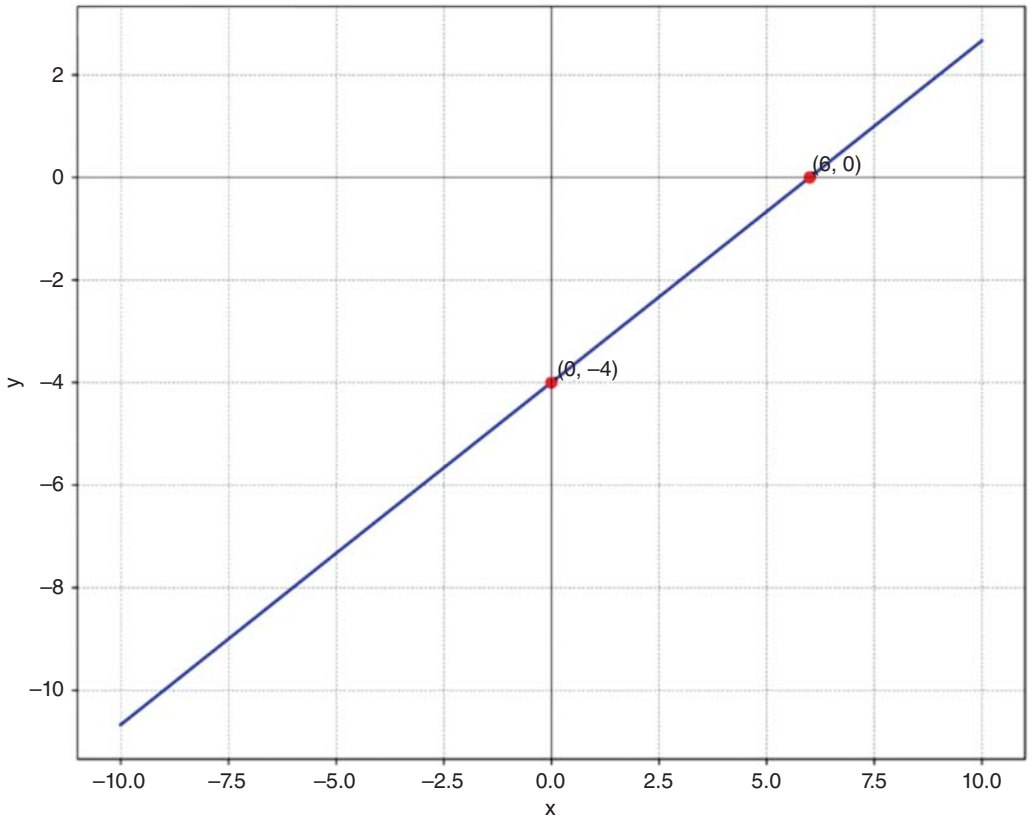
$2(0) - 3y = 12$
$-3y = 12$
$y = -4$

The y-intercept is $(0, -4)$.

Plot the x-intercept $(6, 0)$ and the y-intercept $(0, -4)$ on the coordinate plane.

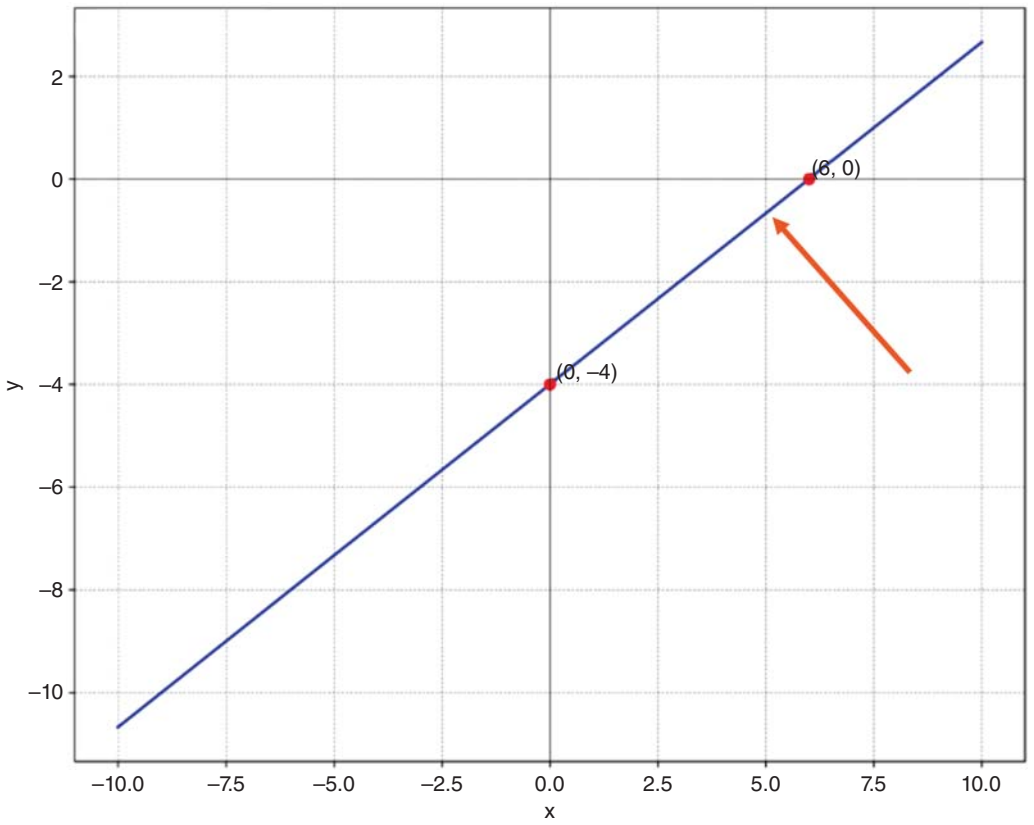
Draw the line passing through these two points.

The resulting graph is:



The line $2x - 3y = 12$ with x-intercept $(6, 0)$ and y-intercept $(0, -4)$.

As mentioned earlier, it is useful to find a third point on the line as a check for accuracy. In the case of $2x - 3y = 12$, any value not coinciding with one of the already-found intercepts will suffice, for either letter. If we use, for instance, $x = 5$, the equation becomes $10 - 3y = 12$, or $-3y = 2$, or $y = -2/3$. The point $(5, -2/3)$ should be on the graph, unless we have made a mistake.



Though we are approximating, note the point appears on the line. Had the point (possibly found with an approximation) is not on the straight-line path, an error has been made, and one should review the steps taken. Unfortunately, it is impossible to immediately tell where the mistake occurred: was it when we found the intercepts, or when we found the third point?

Only a direct review of our work will make this clear.

Example 4.10 Sketch the graph of the line having the equation $4x + 3y = 8$.

Solution: To find the x-intercept, we assign the value of 0 to y and solve for x. This gives:

$4x + 0 = 8$
$4x = 8$
$x = 2$

The x-intercept is the point $(2, 0)$.

To find the y-intercept, we assign the value of 0 to x and solve for y. This gives:

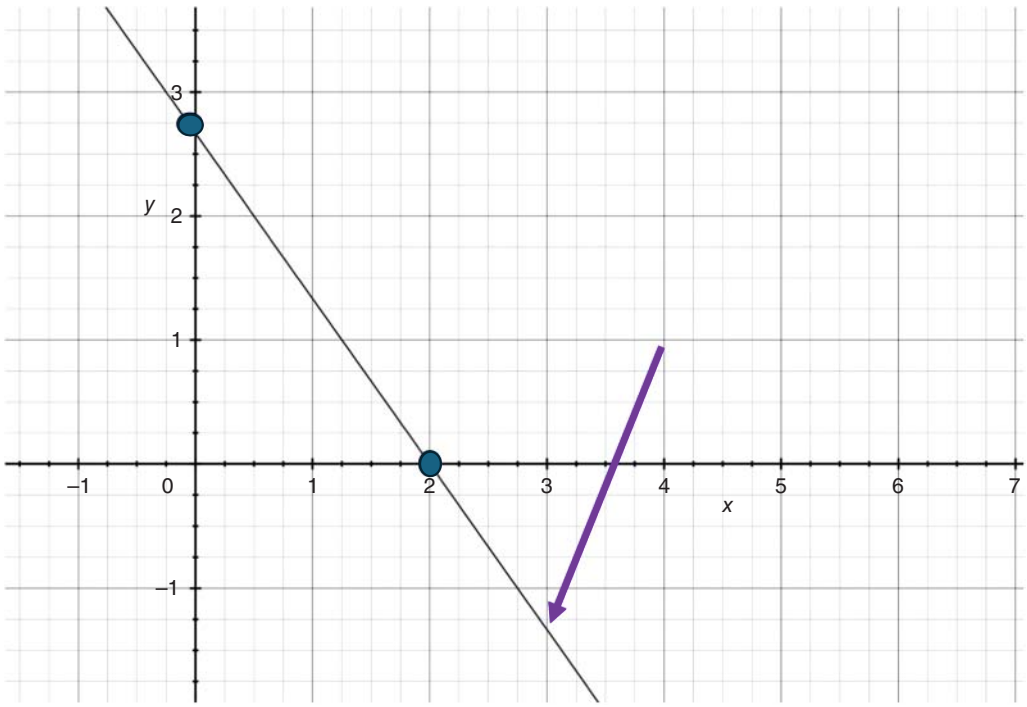
$0 + 3y = 8$
$3y = 8$
$y = 8/3$

The y-intercept is the point $(0, 8/3)$.

A third point on the line, which we can use as a check, can be found by picking some other value for either x or y. Arbitrarily, we will choose $x = 3$; substitution into the equation yields:

$4(3) + 3y = 8$
$12 + 3y = 8$
$3y = -4$
$y = -4/3$

The point $(3, -4/3)$ should be on the line.



Observe the approximate location of the third point bears out our locations of the intercepts, and we are confident in the correctness of our graph.

It is worth noting (in the definition of "Standard Form") to consider the statement "A is non-negative." This does not, we recall, indicate the value of A is "positive." When we say, "A is non-negative," we mean A is either positive or equal to 0. The value of 0 is neither positive nor negative. Observe the subtlety of language in mathematics: there is a difference between saying A is positive, versus saying A is non-negative. These may appear to be the same notions, but they are not!

4.10 Parallel and Perpendicular Lines

A pair of lines in the Cartesian plane is said to be parallel if they maintain a constant distance from one another. In the extreme case, if that distance were zero, the lines would coincide. This tells us every line is, by default, parallel to itself. Since the slope, m, describes the amount the line is tilted from horizontal, we can see this is only possible if the lines possess the same amount of tilt.

Two straight lines are said to be parallel if they have the same slope. Mathematically, if two lines have equations $y = \mathrm{m1}x + \mathrm{b1}$ and $y = \mathrm{m2}x + \mathrm{b2}$, then they are parallel if and only if $\mathrm{m1} = \mathrm{m2}$.

Conversely, two straight lines are said to be perpendicular if the product of their slopes is -1. Mathematically, if two lines have equations $y = \mathrm{m1}x + \mathrm{b1}$ and $y = \mathrm{m2}x + \mathrm{b2}$, then they are perpendicular if and only if $\mathrm{m1} * \mathrm{m2} = -1$.

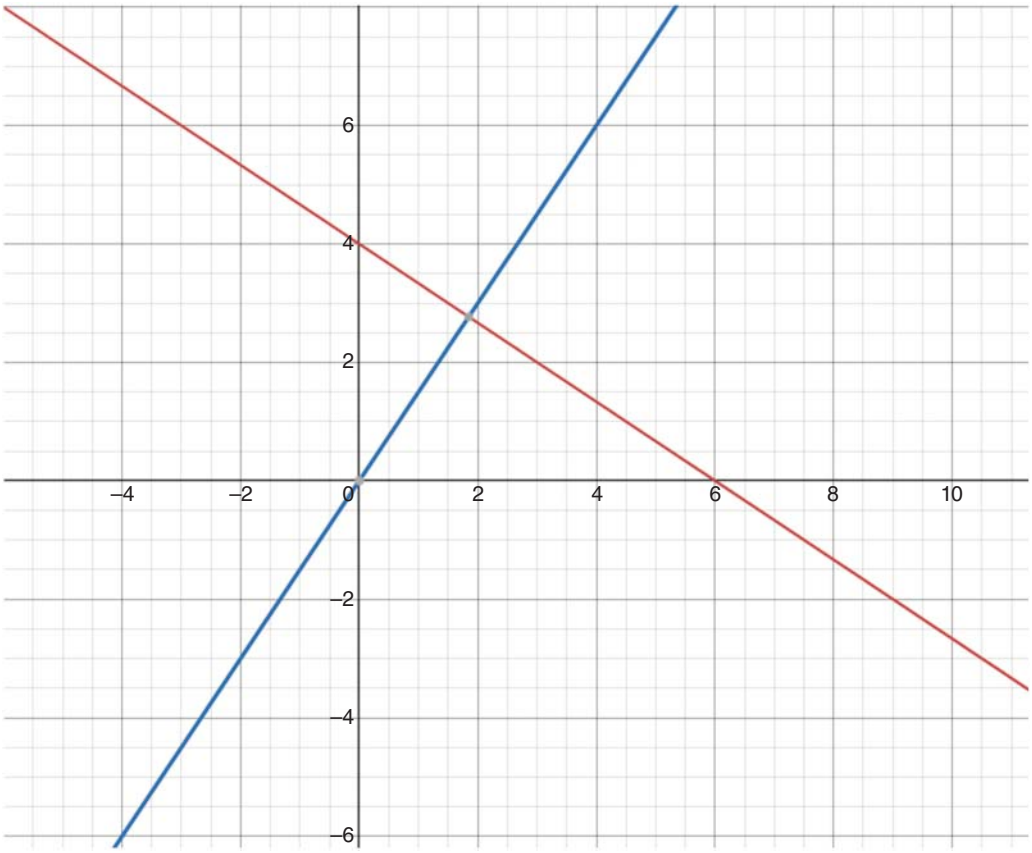
For example, the lines $y = 3x + 2$ and $y = -1/3x + 5$ are perpendicular, as the product of their slopes is -1 ($3 * [-1/3] = -1$).

Another way to describe perpendicular lines is that their slopes are "negative reciprocals" of one another. Reciprocals are fractions that are inverted (or, upside-down) versions of one another— an example of a reciprocal pair is:

$$\frac{2}{7} \text{ and } -\frac{7}{2}$$

Observe this agrees with the first description given, since the product of negative reciprocals is -1.

Below, we graph the two equations on the same plane. For those familiar with concepts of trigonometry, we point out the lines intersect one another at a right angle, or a 90° angle.



4.11 Applications of Straight-Line Equations: Modeling Linear Relationships

Many real-world phenomena can be approximated by linear models, which can provide valuable insights and help make predictions.

The slope–intercept form of a linear relationship can be expressed as:

$$y = mx + b$$

where

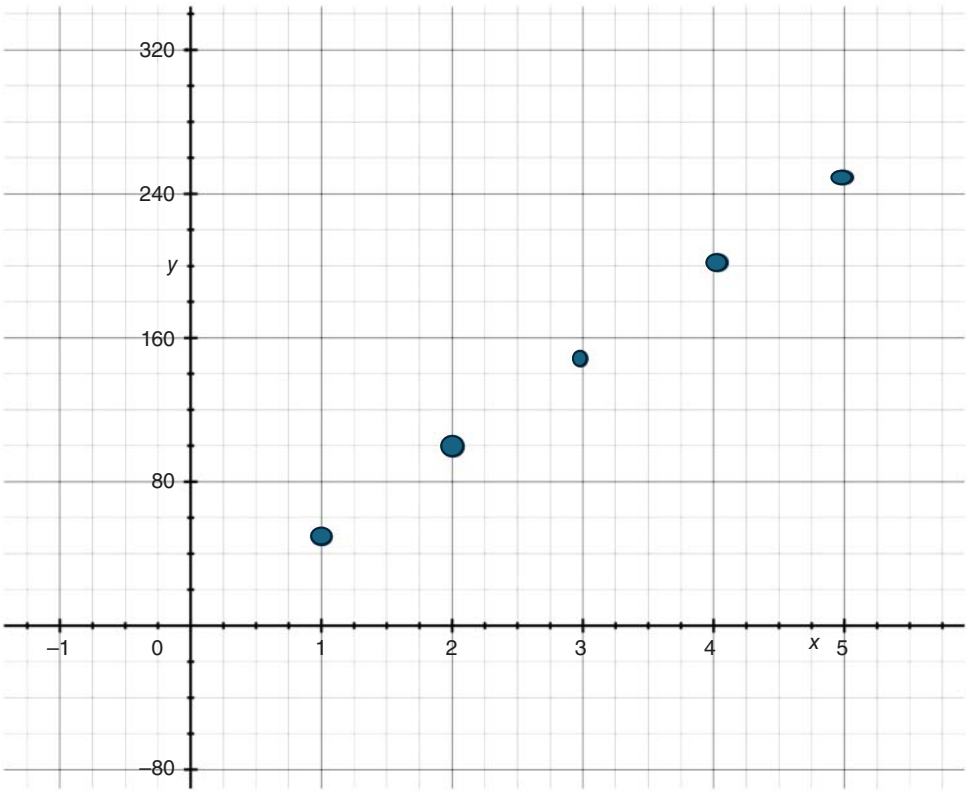
y is the dependent variable;
x is the independent variable;
m is the slope, which represents the rate of change between x and y;
b is the y-intercept, which represents the starting value of y when $x = 0$.

Example 4.11 Modeling the Relationship between Time and Distance

Suppose you are studying the relationship between the time it takes to travel a certain distance and the actual distance traveled. You collect the following data:

Time (t, in hours): 1, 2, 3, 4, 5.
Distance (d, in miles): 50, 100, 150, 200, 250.
Show, by point-plotting, the data appears to follow a linear pattern. Using two of the points, determine the equation of the line apparently being followed, and then show the other points actually lie on the line whose equation you have found. Interpret the slope as a rate of change.

Solution: To model this linear relationship, we can plot the data points on a coordinate plane and fit a straight line to the data using the slope–intercept form:



Note the points appear to fall on a straight-line path. We can use any two of the points to determine the equation of the line and then test the other points to confirm their coordinates satisfy the equation (and so, the points fall on the line).

We will use the points $(t_1, y_1) = (1, 50)$ and $(t_2, y_2) = (2, 100)$. The slope of the line passing through those points is $\mathrm{m} = \frac{100-50}{2-1} = \frac{50}{1} = 50$. Since the vertical axis represents the distance in miles, and the horizontal axis represents time in hours, the units of slope will be miles per hour.

Thus, we have the slope representing 50 miles per hour.

The equation of the line is found using $y - y_1 = m(t - t_1)$, which here would be:

$y - 50 = 50(t - 1)$
$y - 50 = 50t - 50$
$y = 50t.$

One by one, we confirm the other points satisfy this equation. Is, for example, (3, 150) on the graph of that equation?

$y = 50(3) = 150$, so yes, that point is on the line. Similar computations confirm the other points are on the graph of the line, as well.

In the slope–intercept form, the equation of the line is $y = 50t + 0$, and this tells us several things. For every hour of travel, the distance increases by 50 miles. The y-intercept of 0 indicates that the starting distance is 0 miles.

By using this linear model, you can make predictions about the distance traveled given the time, or vice versa. For example, if you travel for 6.5 hours, you can expect to cover a distance of 325 miles (6.5 hours × 50 miles/hour).

4.11.1 Application: Direct Variation (Also Called Direct Proportionality) and Straight-Line Equations

Variation is a situation where you have two variable quantities (things that can change as time goes along) that are somehow connected—like the amount of sunshine in the sky and the temperature of the atmosphere are related. The amount of money we deposit in an account monthly and the future value of the investment are related. There is somehow a connection between the quantities, which we may or may not understand.

While there are several types of variation, we will present "Direct Variation" here.

When we have two quantities that are related through direct variation, they behave the same way as one another: if one increases, so does the other, and if one decreases, so does the other. The amount of change in one quantity is constant per unit change of the other.

The way we will be drawn into the problem will be through a statement of this form:

"A varies directly as B," or "A is directly proportional to B." When we see this, we interpret it mathematically as $\mathrm{A} = k\mathrm{B}$.

Example 4.12 A restaurant owner knows the amount of money taken in at the cashier varies directly with the number of hours the restaurant is open. If it is known the restaurant takes in \$9000 over a span of 4 hours, write a direct variation equation representing the situation and then estimate the amount of time required for the restaurant to bring in \$20,250.

Solution: We begin problems such as this by setting up a legend (a "legend" is a statement in which you identify the variables you are going to use and specify what each of those variables will represent). It proves convenient to choose letters that tie in directly to our concepts, so we do not forget which variable is associated with which quantity.

Here, let us identify M = the money the restaurant takes in, and H = hours the restaurant is open. You could use any variables that make sense to you, or you could always just use A and B if you prefer. So long as a legend is present, you can always identify what each variable represents.

The variation/proportionality equation is constructed from the following relationship: M varies directly as H, so $M = kH$. This is the "variation equation." Note that we have NOT yet used any of the numerical values in the problem! This is an important thing to be aware of when working on a proportionality problem.

Once the equation is set up, we begin to work with the values given in the problem statement. The value of "k" is not known to us, but we will use the known situation to determine the value of k.

We are told the restaurant generates \$9000 of income over a span of 4 hours. Since M represents the money and H the hours, this says.

$9000 = k(4)$.

We solve for k by dividing each side of the equation by 4, obtaining $2250 = k$.

When we have used the set of known measurements to establish the value of k, we go to our variation equation (remember, that was $M = kH$), and we replace k by its now-known value. Here, our equation becomes $M = 2250H$.

The given information has now served its purpose, so we do not use it again. We next turn to our question: **How long would it take for the restaurant to bring in \$20,250?**

That value is a money amount, so it's an M-value. We plug it into our variation equation:

$$20{,}250 = 2250H$$

To solve, we divide each side of the equation by 2250 to isolate H, and we get $9 = H$.

The conclusion is that it will take 9 hours for the restaurant to take in \$20,250.

Read that example carefully—it contains all the information we need in our direct variation problems!

Example 4.13 The distance traveled by a ball rolling down a slope varies directly as the square of the time the ball has been rolling. If the ball traveled 27 ft over a span of 3 seconds, what is the distance the ball traveled over a span of 2 seconds?

Solution: We have two unknowns here: D = the distance the ball has rolled, and T = the time the ball has rolled. We are told the distance rolled varies directly as the **square** of the time it's been rolling, so $D = kT^2$.

We use the known measurement (27 ft over a span of 3 seconds) to find the value of k:

$27 = k(9)$
$3 = k$

After replacing k by its established value, our equation becomes $D = 3T^2$. We use this to answer our posed question: **how far did the ball travel over the period of 2 seconds?**

$D = 3(2)^2$
$D = 3(4)$
$D = 12$

So, the ball rolled 12 ft during the 2 second span of time.

5

Introduction to Systems of Linear Equations

5.1 Definition and Notation of Systems of Linear Equations

A system of linear equations is a collection of two or more linear equations that share the same set of variables. In other words, a system of linear equations is a set of equations where each equation describes a linear relationship between the same set of variables.

Mathematically, a system of n linear equations in m variables can be expressed as:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n \end{aligned}$$

where a_{ij} and b_i are real numbers (some possibly zero), and $x_1, x_2, \ldots, x_m$ are the variables.

The values a_{ij} represent the weights or coefficients of the variables, and the constants b_i are isolated on the right-hand side of each equation.

When we say there is a "linear relationship" among the variables, we mean the equations are linear equations: each of the variables appears to the power 1 only and is not multiplied against other variables. In a graphical sense, the equations represent lines only in the two-dimensional case. When more than two variables appear, the equations have graphs that lie in higher-dimensional space and that are not straight lines. For example, an equation such as $2x_1 + 7x_2 - 5x_3 = 10$ represents a plane in three-dimensional space, despite the fact we refer it as a linear equation in three variables.

Consider the following system of two linear equations in two variables:

$$\begin{aligned} 2x_1 + 3x_2 &= 10 \\ x_1 - x_2 &= 3 \end{aligned}$$

Here, we have a system of two linear equations in the variables: x_1 and x_2. It is often the case that (when two variables are present) we use the letters x and y in place of subscripted variables, and that when three variables are present, we use the letters x, y, and z.

5.2 Graphical Representation of a System of Linear Equations

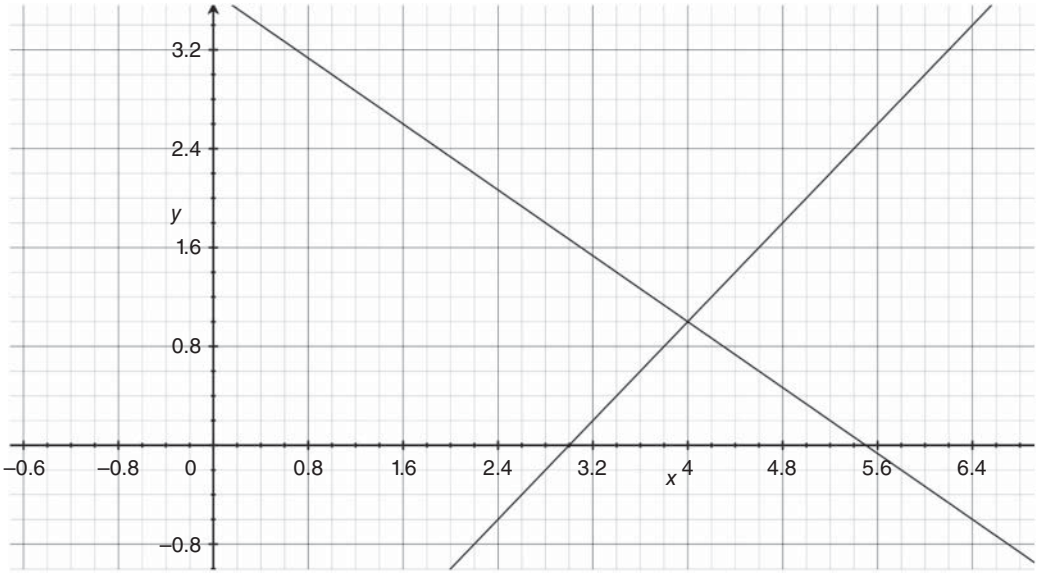
A system of linear equations in two variables can be graphically represented in the coordinate plane. Each equation in the system can be represented as a line

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

Example 5.1 Graphically represent the following system of two linear equations in two variables:

$$2x_1 + 3x_2 = 11$$
$$x_1 - x_2 = 3$$

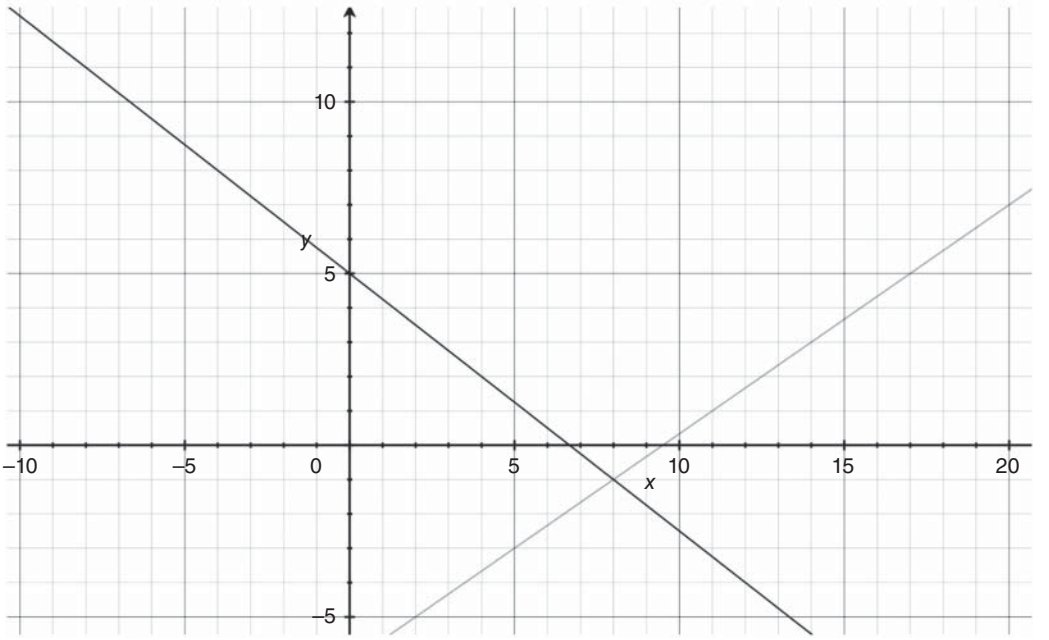
When graphed on the same coordinate plane, we find:



The point of intersection of the two lines (if any) is a key point of interest when considering a system of equations, which we see here is (4,1). Here, the point of intersection happens to have coordinates rather easily determined from an inspection of the graph. As one can imagine, this may not be the case, and so computational methods that avoid reliance upon graphs prove essential.

Example 5.2 Graphically represent the system of two linear equations in two variables.

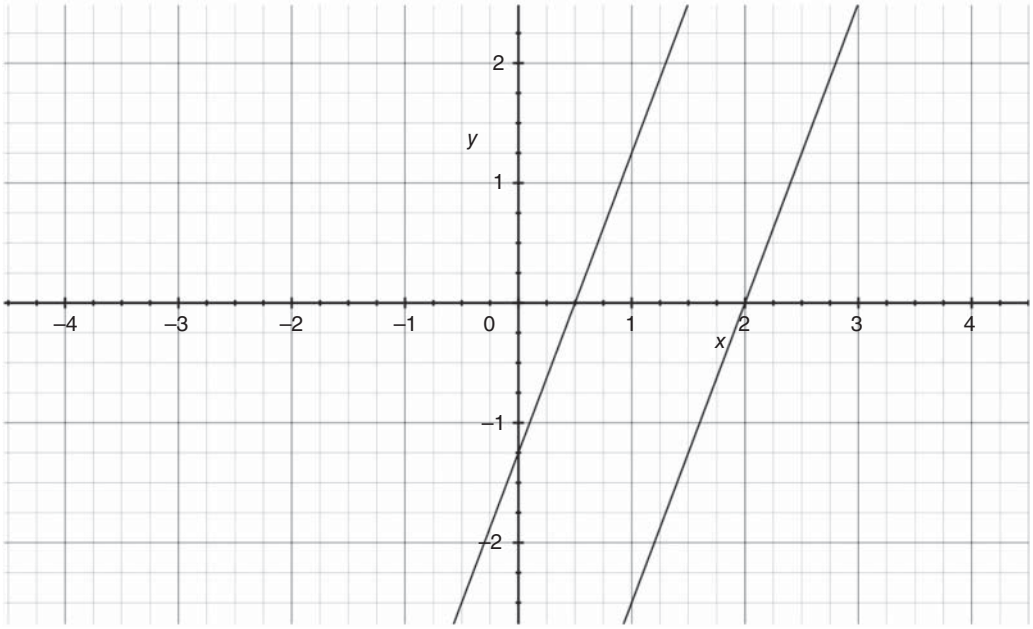
$$2x_1 - 3x_2 = 19$$
$$3x_1 + 4x_2 = 20$$



Observe the lines intersect at the point $(8, -1)$, which is the solution to the system.

Example 5.3 Graphically represent the system of two linear equations in two variables.

$$5x_1 - 2x_2 = 10$$
$$10x_1 - 4x_2 = 5$$



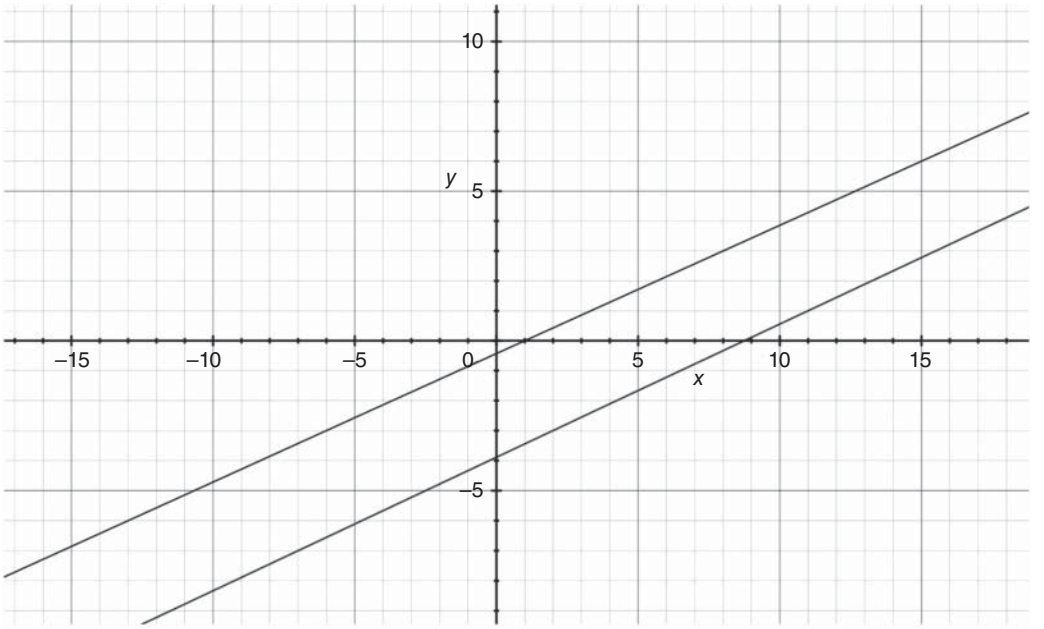
Here we see the two lines of the system are parallel to one another, and therefore have no intersection.

> Note: It might be more cautiously stated the lines in the system "appear" to be parallel. Using graphing technology, it may seem two lines are parallel (even if they are not).

Though we are asserting this without proof, we can verify the lines are parallel by comparing the slopes of each and showing those slopes to be equal. This is the defining quality of parallel lines in the Cartesian plane.

Example 5.4 Graphically represent the system of two linear equations in two variables and recognize they appear to be parallel.

$$12x_1 - 28x_2 = 12$$
$$4x_1 - 9x_2 = 35$$



Superficial examination of the graph suggests the lines are parallel; you should verify they are not!

5.3 The Solution to a System of Two Linear Equations in Two Variables

The point of intersection, as mentioned a moment ago, is of particular interest when regarding the graphical representation of a system of linear equations. This point is referred to as the "solution" to the system.

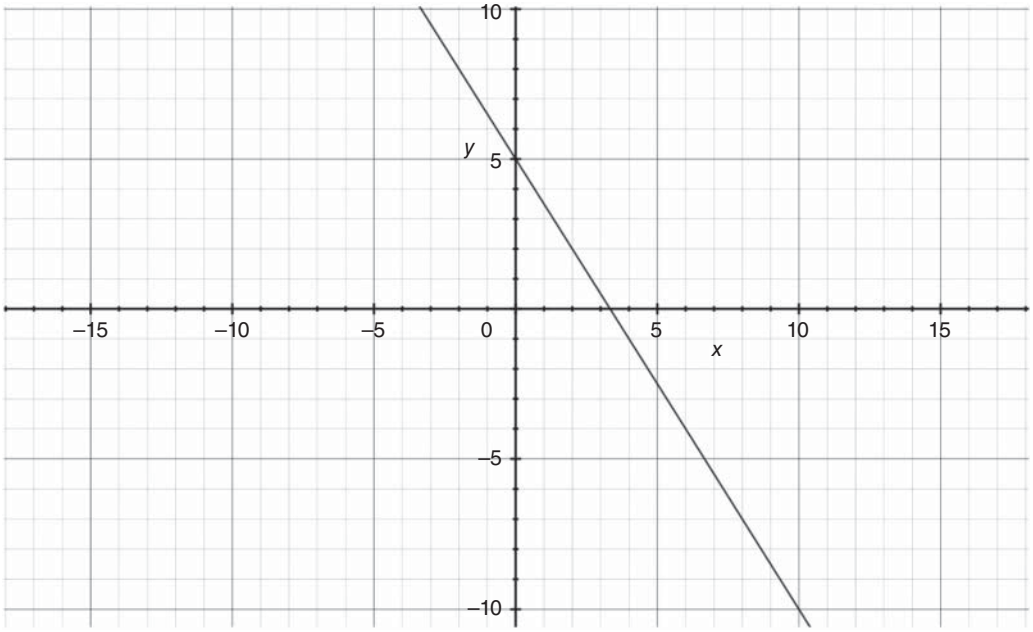
Systems of linear equations may have a single solution; we refer to such systems as "consistent." Systems may, on the other hand, have no solution; we refer to such systems as "inconsistent." As we have seen, inconsistent systems of two linear equations in two variables are inconsistent precisely when their graphical representation reveals parallel lines.

A third possibility exists when considering a system of two linear equations in two variables, which may not immediately occur to you when imagining such a system. That is, our intuition may tell us that two lines either intersect (their graphical representation is a system that is consistent) or do not intersect (their graphical representation is a system that is inconsistent). However, the two equations may have a graphical representation in which the two lines "coincide." This means the graphs of the equations are precisely the same. This occurs when one equation is an exact multiple of the other.

This final possibility results in a situation in which we say the system is "consistent and dependent." The terminology differs from that of the first possible outcome in that "dependent" indicates we have a consistent system that has an endless or infinite number of solutions, since the lines intersect at every point on their mutual graph.

Example 5.5 Graphically represent the system of two linear equations in two variables.

$$3x_1 + 2x_2 = 10$$
$$-6x_1 - 4x_2 = -20$$



Observe the result is a pair of lines in the plane that "overlay" one another.

5.3.1 Checking the Solution to a System of Two Linear Equations in Two Variables

Points on the graph of a line must "satisfy" the equation being graphed. That is, if the coordinates of the points are substituted for the variables in the equation, what results must be a true equality. A proposed solution to a system of linear equations can be checked by ensuring the coordinates satisfy all the equations of the system.

In the system

$$2x_1 - 3x_2 = 19$$
$$3x_1 + 4x_2 = 20$$

we found the solution was the point $(8, -1)$. This result is checked by verifying the coordinates produce true equalities in both equations of the system.

$$2(8) - 3(-1) = 19$$
$$16 - (-3) = 19 \text{ true}$$
$$3(8) + 4(-1) = 20$$
$$24 + (-4) = 20 \text{ true}$$

5.3.2 Statement of the Solution to Linear Systems of Equations

As stated, a system of linear equations can be classified as consistent, inconsistent, or dependent. Let us remind ourselves of the terminology and indicate how the solutions should be presented.

5.3.2.1 Consistent System

A system of linear equations is consistent if it has at least one solution, that is, there exists at least one set of values for the variables that satisfies all the equations in the system.

In the graphical representation, a consistent system of two linear equations in two variables has a unique point of intersection, which represents the solution to the system. The solution is stated as an ordered pair, given within the braces that represent membership in a set: $\{(x_1, x_2)\}$.

5.3.2.2 Inconsistent System

A system of linear equations is inconsistent if it has no solution, that is, there is no set of values for the variables that satisfies all the equations in the system. A mathematical way of expressing this is that "the set of solutions to the system is empty." Recall our notations for the empty set were $\emptyset$ and $\{\}$. We can use either of those notations, or we can use the phrase "no solution" to describe the solution to the system to indicate there is no point of intersection (the lines are parallel).

5.3.2.3 Dependent System

It is possible the two equations could have the same graph in the plane. We say, in such a situation, the lines "coincide." The solution to such a system will be given in what we call "set-builder notation,"

$$\{(x_1, x_2) \mid \text{either equation of the system}\}.$$

We can use either equation because they are equivalent forms of the same equation, and thus have the same graph.

Example 5.6

$$4x_1 - 5x_2 = 13$$
$$-8x_1 + 10x_2 = -26$$

If we examine the equations of the system, we observe the two equations are scalar multiples of one another, which was a hallmark of the dependent system. We would give either of the following as our solution:

Solution:

$$\{(x_1,\ x_2) \mid 4x_1 - 5x_2 = 13\}$$

or

$$\{(x_1,\ x_2) \mid -8x_1 + 10x_2 = -26\}$$

5.3.3 Solving Systems of Linear Equations by the Substitution Method

To this point, the solutions to the systems have been found graphically, recognizing the point of intersection of two linear graphs. This becomes challenging (or impossible) when the point of intersection (if any) has coordinates that are not integer-valued. For this reason, we must develop computational techniques that are immune to this weakness.

When dealing with systems of linear equations, the substitution method is a powerful computational technique for finding the solution. This method involves isolating one variable in one of the equations and then substituting the expression to which it is found to be equal into the other equations to eliminate that variable.

In what follows, it suits us to use the variables x and y, as these are the variable choices with which we are most familiar.

The general steps for the method of substitution are:

1) Select one of the equations in the system with which to work.
2) In that equation, select one letter to algebraically isolate, and do so.
3) In the other equation, substitute for the letter selected for isolation in step 2 with the expression it was found to be equal.
4) Solve the resulting single equation.
5) Return to the equation from step 2 (where the selected variable was isolated) and evaluate using the result from step 4.
6) State the solution to the system.

Let us walk through an example to illustrate the process.

Consider the following system of linear equations:

$$3x + 2y = 12$$
$$x - y = 1$$

Step 1: To solve this system using the substitution method, we will first select an equation; here, we will work with the second equation (it appears to be simpler than the first).

Step 2: Next, we choose one of the variables and work to isolate it within its equation. Arbitrarily, we can elect to isolate the variable y:

$$x - y = 1$$
$$-y = 1 - x$$
$$(-1)(-y) = (-1)(1 - x)$$
$$y = -1 + x$$
$$y = x - 1$$

Step 3: Now, we can substitute this expression for y into the first equation of the system (the equation we did not select earlier):

$$3x + 2(x - 1) = 12$$

Step 4: Solve this equation:

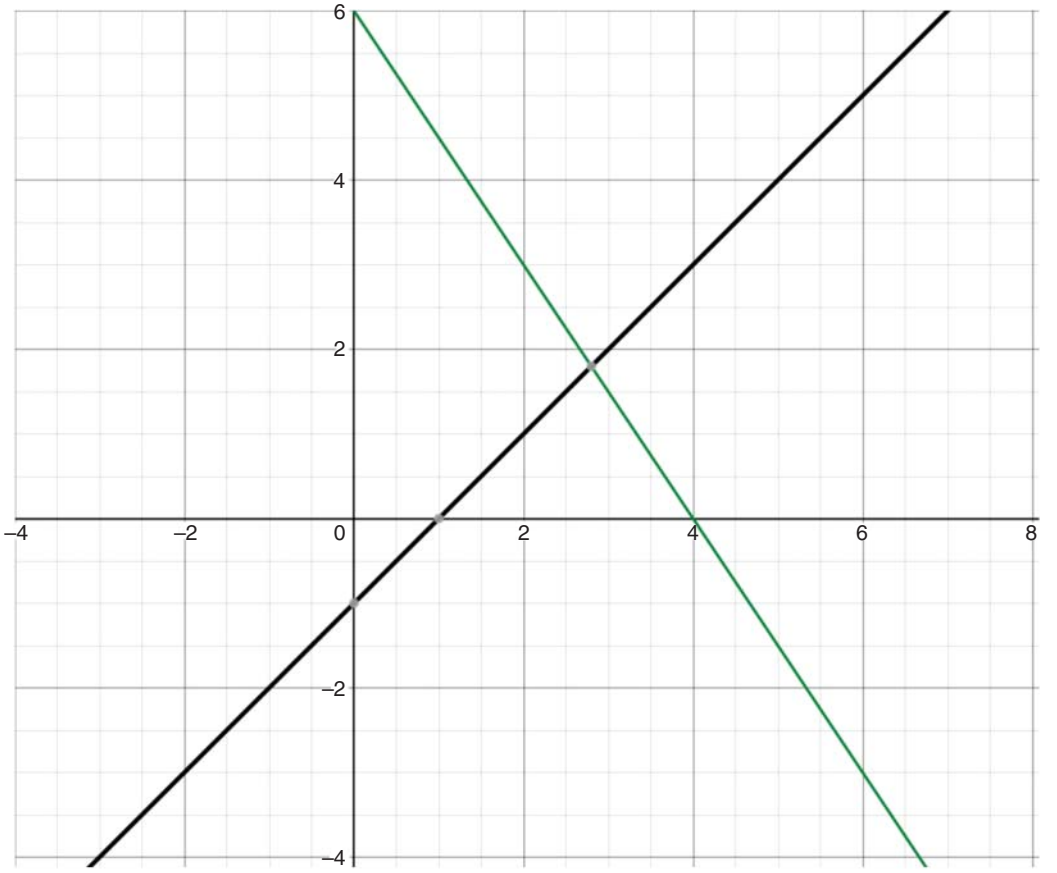
$$3x + 2x - 2 = 12$$
$$5x = 14$$
$$x = 2.8$$

Step 5: Substituting the value of x back into the form of the second equation where y was isolated, we can find the value of y:

$$y = x - 1$$
$$y = 2.8 - 1 = 1.8$$

Step 6: Therefore, the solution to the system of linear equations is $x = 2.8$ and $y = 1.8$.

Here is a graphical representation of the system:



Although a visual determination of the solution is challenging, we can see it does appear the two lines intersect at the point (2.8, 1.8), and therefore the graph lends support to the claim that {(2.8, 1.8)} is the solution to the system.

> Note 1: We see the importance of having a computational method for finding the solution to a system of equations: the exact point of intersection may be challenging to identify if the coordinates do not have integer values.
>
> Note 2: The choices made in the substitution method are arbitrary and do not affect the outcome of the technique. Let us solve the system again, this time making a different choice.

Example 5.7 ***Revisited:***

$$\begin{aligned} 3x + 2y &= 12 \\ x - y &= 1 \end{aligned}$$

Solution:

Step 1: This time, we will choose to work with the first equation in the system.

Step 2: In the first equation, we choose to solve for the variable x.

$$3x + 2y = 12$$
$$3x = 12 - 2y$$
$$x = 4 - (2/3)y$$

Step 3: We turn to the second equation and substitute for x,

$$4 - (2/3)Y - Y = 1$$

Step 4: We solve this equation:

$$4 - (5/3)Y = 1$$
$$-(5/3)y = -3$$
$$y = (-3)(-3/5)$$
$$y = 9/5$$
$$y = 1.8$$

Step 5: Returning to $x = 4 - (2/3)y$ and substituting this value of y, we find:

$$x = 4 - (2/3)(9/5)$$
$$x = 4 - 18/15$$
$$x = 4 - 6/5$$
$$x = 20/5 - 6/5$$
$$x = 14/5$$
$$x = 2.8$$

Step 6: Again, we have shown the solution to be $\{(2.8, 1.8)\}$.

Example 5.8 Solve the following system of linear equations:

$$2x + 3y = 12$$
$$-x + y = 5$$

Solution:
Step 1: We will choose to work with the second equation.
Step 2: In this equation, we will choose to isolate the variable y.

$$-x + y = 5$$
$$y = x + 5$$

Step 3: Substitute the expression for y into the first equation.

$$2x + 3(x + 5) = 12$$
$$2x + 3x + 15 = 12$$

Step 4: Solve for x.

$$5x + 15 = 12$$
$$5x = -3$$
$$x = -3/5$$

Step 4: Substitute the value of x into the second equation.

$$y = x + 5$$
$$y = -3/5 + 5$$

Step 5: Calculate the value of y.

$$y = (-3/5) + 5$$
$$y = -3/5 + 25/5$$
$$y = 22/5.$$

Step 6: State the solution to the system:
The solution to the system is $\{-3/5, 22/5)\}$.

Example 5.9 Solve the following system of linear equations:

$$x + y = 4$$
$$3x - 2y = 7$$

Solution:
Step 1: We will choose to work with the first equation of the system.
Step 2: Solve the first equation for one of the variables, in this case y.

$$x + y = 4$$
$$y = 4 - x$$

Step 3: Substitute the expression for y into the other equation and solve.

$$3x - 2(4 - x) = 7$$
$$3x - 8 + 2x = 7$$
$$5x = 15$$
$$x = 3$$

Step 4: Substitute the value of x into the first equation.

$$y = 4 - x$$

Step 5: Simplify to find the value of y.

$$y = 4 - 3$$
$$y = 1$$

Step 6: Therefore, the solution to the system is $\{(3, 1)\}$.

Example 5.10 Solve the following system of linear equations:

$$2x - 3y = 11$$
$$4x + y = 14$$

Solution:
Step 1: We will choose to work with the second equation

Step 2: Solve for one of the variables, in this case x.

$$4x + y = 14$$
$$x = (14 - y)/4$$

Step 3: Substitute the expression for x into the first equation and solve.

$$2((14 - y)/4) - 3y = 11$$
$$28 - 2y - 12y = 44$$
$$-14y = 16$$
$$y = -4/7$$

Step 4: Substitute the value of y into the second equation.

$$x = (14 - y)/4$$

Step 5: Simplify to find the value of x.

$$x = (14 - [-4/7])/4$$
$$x = 18/7/4$$
$$x = 9/7$$

Step 6: Therefore, the solution to the system is $\{(9/7, -4/7)\}$.

> Note: Unless specifically told to do otherwise, we are attempting to find the exact solution to the system. It is incorrect to use a decimal approximation (no matter how fine), unless specifically told to do so.

5.3.4 Substitution, the Inconsistent Case, and the Dependent Case

How would we recognize, when using the substitution method, a system of linear equations was inconsistent? What about a system that is dependent? We know such outcomes are possible, so it is crucial we be able to recognize them when they appear.

A good way to think of the inconsistent case is that it is "impossible to find a solution." Mathematically, this would suggest we would encounter a statement that is a contradiction (a statement that is patently false or absurd).

Having familiarized ourselves with the step-by-step process involved in substitution, we will omit references to the specific steps to streamline the work.

Example 5.11 Solve the system of linear equations in two variables using substitution.

$$4x + y = 11$$
$$8x + 2y = 12$$

We begin by choosing to work with the first equation and isolate the variable y.

$$y = 11 - 4x$$

Turning to the second equation and performing a substitution for y,

$$8x + 2(11 - 4x) = 12$$

Attempting to solve, we find:

$$8x + 22 - 8x = 12$$

$$22 = 12$$

This statement is a contradiction; this is where we should recognize the system of equations is impossible to solve and declare there is no solution to the system.

A dependent system would likewise be recognized by the appearance of an extreme computational result. In this case, the computations would lead to an identity statement, which is true for all values of the variable involved.

Example 5.12 Solve the system of linear equations in two variables using substitution.

$$x + 5y = 6$$

$$2x + 10y = 12$$

As before, we choose an equation with which to work, and within that equation a variable to isolate. Here, it is convenient to choose the first equation and isolate x.

$$x + 5y = 6$$

$$x = 6 - 5y$$

Turning to the first equation, we substitute for x:

$$2(6 - 5y) + 10y = 12$$

$$12 - 10y + 10y = 12$$

$$12 = 12$$

We have reached a line of work that is an identity statement. This indicates the system is dependent, and we identify the solution as $\{(x, y) | \, x + 5y = 6\}$.

5.3.5 Solving Systems of Equations Using the Method of Elimination

A second computational technique for solving systems of linear equations is the method of elimination (referred to by some as the "addition method"). This approach involves manipulating the equations in the system in such a way as to eliminate one of the variables, allowing you to solve for the remaining variable. Once you have solved for one variable, you can then substitute that value back into one of the original equations to solve for the other variable.

The general steps for the method of elimination are:

1) Multiply one or both of the equations by a constant to make the coefficients of one variable have the same numerical value in both equations yet having opposite signs.
2) Add the two equations to eliminate one of the variables.
3) Solve the resulting single equation for the remaining variable.
4) Substitute the value of the variable you just solved for back into one of the original equations to find the value of the other variable.

Let us walk through an example to illustrate the process:

Example 5.13 Solve the following system of linear equations using the method of elimination:

$$2x + 3y = 12$$

$$x - y = 5$$

Step 1: Multiply the second equation by 3 to eliminate variable y.

$3(x - y = 5)$

$3x - 3y = 15$

Step 2: Add the two equations to eliminate the y variable.

$2x + 3y = 12$

$3x - 3y = 15$

$5x = 27$

Step 3: Solve for x.

$5x = 27$

$x = 27/5$

Step 4: Substitute the value of x into either of the original equations to find the value of y.
We shall use the second equation: $x - y = 5$

$27/5 - y = 5$

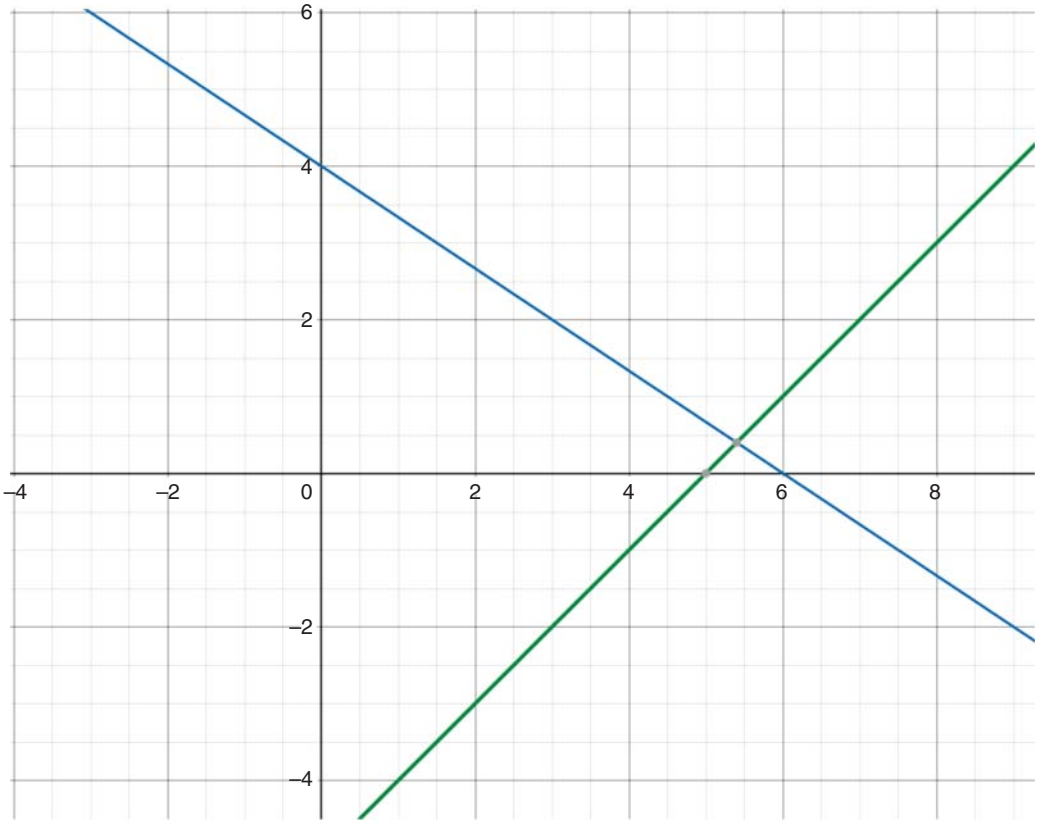
$-y = 5 - 27/5$

$-y = 25/5 - 27/5$

$-y = -2/5$

$y = 2/5$

Therefore, the solution to the system of linear equations is $\{(27/5, 2/5)\}$.
Here is a visual representation of the solution:



This graph shows the two lines representing the system of equations, with the point of intersection visually estimated at (5.4, 0.4), which is the solution in decimal form.

Example 5.14 Solve the following system of equations:

$$3x + 2y = 11$$
$$x - y = 1$$

Solution:
Step 1: Multiply the second equation by 2 to have the coefficients of y agree, except for sign.

$$2(x - y = 1)$$
$$2x - 2y = 2$$

Step 2: Add the two equations to eliminate y.

$$3x + 2y = 11$$
$$2x - 2y = 2$$
$$5x = 13$$

Step 3: Solve for x.

$$x = 13/5$$

Step 4: Substitute the value of x into one of the original equations to find y.
Using the first equation:

$$3(13/5) + 2y = 11$$
$$39/5 + 2y = 11$$
$$2y = 11 - 39/5$$
$$2y = 55/5 - 39/5$$
$$2y = 16/5$$
$$y = 16/10$$
$$y = 8/5$$

Therefore, the solution to the system is $\left\{\left(\frac{13}{5}, \frac{8}{5}\right)\right\}$.

Example 5.15 Solve the following system of equations:

$$4x + 3y = 17$$
$$2x - y = 5$$

Solution:
Step 1: Multiply the second equation by 3 to have the coefficients of y be equal, apart from the sign.

$$3(2x - y = 5)$$
$$6x - 3y = 15$$

Step 2: Add the equations to eliminate the variable y.

$$4x + 3y = 17$$
$$6x - 3y = 15$$
$$-2x = 2$$

Step 3: Solve for x.

$$x = -1$$

Step 4: Substitute the value of x into one of the original equations to find y.
Using the second equation:

$$2(-1) - y = 5$$
$$-2 - y = 5$$
$$y = -7$$

Therefore, the solution to the system is $\{(-1, -7)\}$.

Example 5.16 Solve the following system of equations using elimination, rounding your answer to a decimal approximation accurate to the nearest hundredth.

$$2x + 3y = 11$$
$$5x - 2y = 13$$

Solution:
Step 1: Multiply the first equation by 5 and the second equation by -2 to make the coefficients of x equal but opposite in sign.

$$5(2x + 3y = 11)$$
$$10x + 15y = 55$$
$$-2(5x - 2y = 13)$$
$$-10x + 4y = -26$$

Step 2: Add the equations together.

$$10x + 15y = 55$$
$$\underline{-10x \pm 4y = -26}$$
$$19y = 29$$

Step 3: Solve for y.

$$y = 29/19 \approx 1.53$$

Step 4: Substitute the value of y into one of the original equations to find x.
Using the first equation:

$$2x + 3(1.53) \approx 11$$
$$2x + 4.59 \approx 11$$
$$2x \approx 6.41$$
$$x \approx 3.21$$

Therefore, the solution to the system is approximately $\{(3.21, 1.53)\}$.

Example 5.17 Solve the following system using elimination.

As mentioned before, systems can have NO solution. This, keep in mind, would happen if the two lines happened to be parallel, and if the solution process leads us to a mathematical contradiction or absurdity.

$$2x - 3y = 3$$
$$-4x + 6y = 6$$

Step 1: Choose to eliminate one of the variables, in this case x. Multiply the first equation by 2,

$$2(2x - 3y = 3)$$
$$-4x + 6y = 6$$
$$4x - 6y = 6$$
$$-4x + 6y = 6$$

Step 2: Add the equations.

$$4x - 6y = 6$$
$$-4x + 6y = 6$$
$$0 = 12$$

Note the work has led us to a contradiction; this informs us the system is inconsistent, and we can identify the solution as $\emptyset$.

As a final illustration, we present a dependent system of equations and show it would (as was the case in the substitution method) be recognized by the appearance of an identity statement.

Example 5.18 Solve the following system of two linear equations in two variables by elimination.

$$10x + 4y = 18$$
$$5x + 2y = 9$$

We multiply the second equation by -2, to obtain

$$10x + 4y = 18$$
$$-10x - 4y = -18$$

Add the equations together,

$$0 = 0$$

This is an identity, and therefore the system is dependent. Using the first equation of the system, we can represent the solution as $\{(x, y) \mid 10x + 4y = 18\}$.

5.4 Applications That Lead to Systems of Linear Equations

There are situations that arise in many areas of mathematics and science whose nature leads to a system of equations. While numerous illustrations exist, let us consider a particular problem referred to as a mixture problem.

Example 5.19 Mixing Solutions
Another common application of systems of linear equations is mixing solutions of different concentrations.

Suppose we have two solutions, one with a 40% alcohol concentration and the other with a 20% alcohol concentration. We want to mix these solutions to create 10 liters of solution with a 30% alcohol concentration. How much of each solution do we need?

We assign variables to represent the quantities being sought:

x = amount of 40% solution (in liters)
y = amount of 20% solution (in liters)

We can write two linear equations to represent the constraints:
The total volume of the mixed solution is 10 liters:

$$x + y = 10$$

When considering fluids having a partial constituent of alcohol, the amount of alcohol in that fluid is found by multiplying the concentration by the quantity of fluid. That is, for the 40% alcohol solution (of which we have x liters), the amount of alcohol present is $0.4x$ liters. Similarly, the amount of alcohol contained in the 20% alcohol solution (of which we have y liters) is $0.2y$ liters, and the total alcohol content in the mixed solution is 30% of the 10 l, or $(0.3)(10) = 3$ liters.

Now we have a system of two linear equations with two unknowns:

$$x + y = 10$$
$$0.4x + 0.2y = 3$$

We can solve this system using one of the techniques we have seen to this point. Here, we will apply the substitution method:

Solve the first equation for y:

$$y = 10 - x$$

Substitute into the second equation:

$$0.4x + 0.2(10 - x) = 3$$
$$0.4x + 2 - 0.2x = 3$$
$$0.2x = 1$$
$$x = 5$$

Plug $x = 5$ back into the first equation to solve for y:

$$y = 10 - x$$
$$y = 10 - 5$$
$$y = 5$$

Recalling the quantities represented by the variables, the amounts of each solution needed to create the 30% alcohol solution are 5 l of the 40% solution

5 liters of the 20% solution

This example demonstrates how a system of linear equations can be used to solve a real-world problem involving mixing solutions of different concentrations. It is a coincidence the amounts of each constituent solution are precisely equal.

5.4.1 Systems Involving Three Variables

An equation, recall, is referred to as "linear" if the powers on all the variables are exclusively 1, and none of the variables occurs as a denominator or is multiplied by another of the variables.

In the 3-variable case, such equations would have the form

$$Ax + By + Cz = D$$

where the + signs could just as easily be – signs.

In the systems that we are going to be examining, we will restrict our attention to the case where there are as many equations as there are variables; hence, our systems will have three equations and three variables in this section.

An example of such a system would be

$$3x + 4y - z = 17$$
$$5x - y + 2z = -2$$
$$2x - 3y + 7z = -21$$

The solution to a system of this type would be a point (or points) in three-dimensional space that could satisfy all three of the equations SIMULTANEOUSLY. Understandably, it is difficult to imagine that we could find such a point, even if it were to exist.

It is worthwhile, in the case of three variables, to consider the meaning of each of the three "linear" equations that form the system. What do they represent?

Since there are three variables, the graphs of the equations lie in three-dimensional space. The term "linear" is a bit misleading, because it guides our belief toward the notion each of the equations represents a line in that space. This is not true, however.

A linear equation in three variables has a graph which is a plane in three-dimensional space.

We will use a form of the elimination process that was used in the two-dimensional (two equation, two variable) systems that we just examined. The process that we will employ is referred to as Gaussian Elimination.

Gaussian Elimination is a process that permits us to perform three operations, and three operations only, upon the equations of a system. The operations are often done more than once each and are referred to as Elementary Row Operations.

The three Elementary Row Operations are:

- Any two equations may be "interchanged" in position at any time.
- Any equation may be multiplied by any nonzero constant at any time.
- A multiple of any equation may be added to any other equation at any time, with only the equation that is being "added to" changing.

The object of the application of these three operations is to rewrite the given system in an equivalent form, called ROW ECHELON FORM. From that form, the system can easily be solved by a process called BACK SUBSTITUTION.

Informally, "Row Echelon Form" is a situation where the top equation in the system involves x (with a coefficient of 1), while the other two equations do not involve x, the second equation involves y (with a coefficient of 1), while the third equation does not involve y, and the third equation involves only the variable z (with coefficient of 1).

Formally, "Row Echelon Form" is a situation where each equation of the system has a leading coefficient of 1, each equation of the system presents the variables in the same order, and each equation has fewer variables than the equation above it.

Such a system would look like this:

$$x - 3y + 2z = 4$$

$$y + 4z = 9$$

$$z = 2$$

Given any system of linear equations, our objective is to rewrite it in this "Row Echelon Form." How we will do that is, as yet, unclear, but IF we had this form, we could find the solution to the system by the method of "Back Substitution." This starts at the bottom of the system and works upward.

The bottom equation tells us that $z = 2$. If we substitute that value into the second equation, we can solve for y.

$$y + 4(2) = 9$$

$$y + 8 = 9$$

$$y = 1$$

If we take the values of z and y and substitute them into the top equation, we can solve for x, and then we can give our solution to the system as the point (x, y, z), using set-builder notation.

$$x - 3y + 2z = 4$$

$$x - 3(1) + 2(2) = 4$$

$$x - 3 + 4 = 4$$

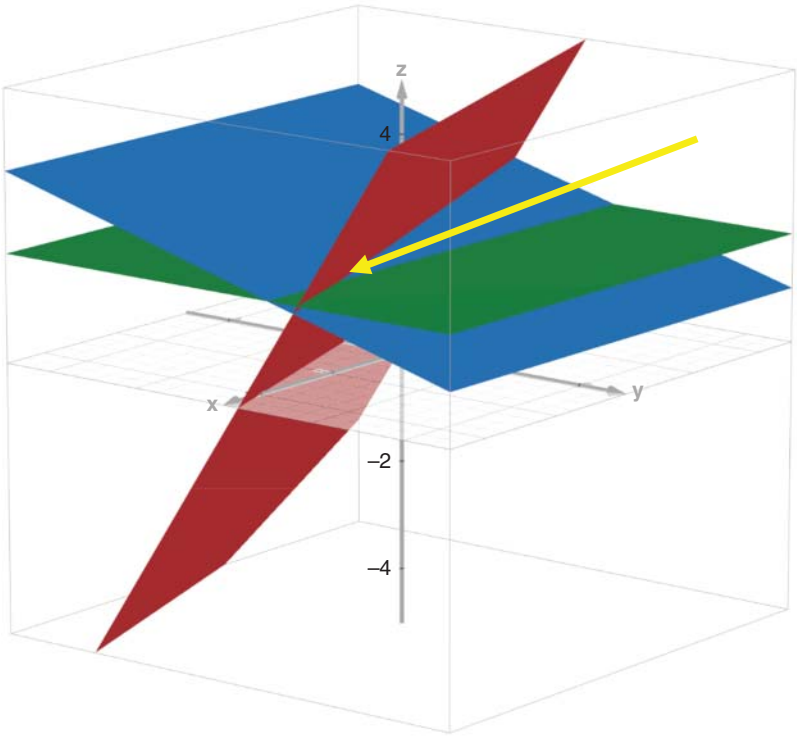
$$x + 1 = 4$$

$$x = 3$$

The solution to the system is {(3,1,2)}.

As before, the solution to the system can be checked by verifying the three coordinates satisfy each of the equations.

Without computer assistance, it is difficult to visualize the situation. However, were we to employ computer software, we would see the three planes (in this case, viewed from the first octant – that part of three-dimensional space where all three coordinate axes are positive) intersecting:



The arrow indicates the solution to the system.

Example 5.20

$$2x + 6y - 4z = -6$$
$$x - y + z = 4$$
$$3x + 2y + z = 5$$

Your first step would be that you would want to start your top equation with just x… note that we have $2x$ there now… there are three possible ways that we could generate x to start the first row:

- divide the entire first equation by 2, every term, and then switch the first two equations
- divide the third equation by 3, and then switch the first and third equations
- (this is less obvious!) add (-1) times the first equation to the third equation, and then switch the first and third equations.

It is important to keep in mind that there is no single correct procedure to achieving Row Echelon Form, and that Row Echelon Form is not "unique." You and another person can do the process differently from one another, obtain different Row Echelon Forms, and both be correct! When you both perform "Back Substitution," you will discover your solution points will have the same set of coordinates.

We will choose to divide the first equation by 2. The result is the system:

$$x + 3y - 2z = -3$$
$$x - y + z = 4$$
$$3x + 2y + z = 5$$

Once you have the first equation starting with x, you must "eliminate" the x's from the lower equations. You do this by adding a multiple of the first equation to the lower equations. The multiple you want will be the NEGATIVE of the existing coefficient of x in the respective equations.

Example 5.21 Remove the x-terms from the two lower equations of the system, using the Gaussian elimination method:

$$x + 3y - 2z = 3$$
$$4x + 8y + z = 18$$
$$-3x - 10y + 7z = 20$$

You would add -4 times the first equation to the second equation, and 3 times the first equation to the third equation… do you see why? The choice is motivated by the existing coefficients of x in the lower equations.

$$x + 3y - 2z = 3$$
$$-4y + 9z = 6$$
$$-1y + z = 29$$

Observe we have removed x (that is, we have eliminated x) from the lower two equations. At this point, the first equation is considered "finished," and we work solely with the two lower equations. The process iterates, and we now attempt to rewrite the system so the coefficient of y in the second equation is 1; we can do this by dividing the second equation by -4, or by dividing the third equation by -1 and then switching the second and third equations.

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$-4y + 9z = 6$$

Next, we eliminate the y-term in the third equation by adding a multiple of the second equation to the third. In this case, we would add four times the second equation to the third, to obtain:

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$5z = -110$$

Next, we divide the final equation by the coefficient of z (here, this is 5), to obtain:

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$z = -22$$

The result is "row echelon form," from which we can solve via back-substitution, which we will do momentarily.

The technique of Gaussian elimination, at the outset, seems Byzantine. Why are we employing such involved manipulations? Is there not a simpler way to achieve our objective of Row Echelon Form?

The answer to the latter equation is that there are alternatives, but none that lend themselves so well to the application of computer technology as a solving aid. The answer to the former question

is that this is an exercise to establish and enhance your ability to perform a series of mental computations and comparisons quickly and accurately. We are entering mathematical provinces in which this ability is of paramount importance, and so it behooves us to work at developing these skills.

Try to think of this process in that way: it is simply a test to see how accurate you can be when doing mental computations. Solving a system of three linear equations by Gaussian Elimination is an excellent way to check your mental acuity, your accuracy, and your attention to detail. Because so much of the work is done "in your head," you must be computationally proficient to get the system to Row Echelon Form correctly.

Once there is an x to start the first equation, and no x's in the lower equations, you then concern yourself with the second and third equations only. These equations form a system of two linear equations in two unknowns and can be treated in a manner similar to the situation of three equations we have just considered.

Shown is the entire computation:

$$x + 3y - 2z = 3$$
$$4x + 8y + z = 18$$
$$-3x - 10y + 7z = 20$$

{Add −4 times row 1 to row 2, and 3 times row 1 to row 3.}

$$x + 3y - 2z = 3$$
$$-4y + 9z = 6$$
$$-y + z = 29$$

{Multiply row 3 by −1, and switch row 2 & row 3.}

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$-4y + 9z = 6$$

{Add 4 times row 2 to row 3.}

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$5z = -110$$

{Divide row 3 by 5}

$$x + 3y - 2z = 3$$
$$y - z = -29$$
$$z = -22 \rightarrow y + 22 = -29$$
$$y = -51$$
$$x + 3(-51) - 2(-22) = 3$$
$$x - 153 + 44 = 3$$
$$x - 109 = 3$$
$$x = 112$$

Thus, the solution is (112, −51, −22).

Example 5.22 Solve the system using Gaussian Elimination

$3x + 4y - z = 17$
$5x - y + 2z = -2$
$2x - 3y + 7z = -21$

Solution:
This system is more complicated than those we have examined to this point, because it is not possible to divide any of the equations by the coefficient of its x-term without producing fractions. While this will certainly work, it will introduce complexity to the mental computations we must perform.

It is conceivable (though it is not one of the steps of Gaussian Elimination) we could rearrange the variables, since interchanging the locations of x and y, for instance, would allow us to have a leading coefficient of 1 for y in the second equation. It is unwise to take this approach, however, as it introduces the additional risk of mis-ordering the coordinates of the solution we find.

Note, leaving the variables in their given order, the coefficients of the x's differ 1 between the first and third equations. This suggests that, if we were to add (-1) times the third equation to the first equation, the result would be that the first equation would have leading x-coefficient of 1.

$3x + 4y - z = 17$

$5x - y + 2z = -2$

$2x - 3y \pm 7z = -21$ Add -1 times eq. 3 to eq. 1 (only the first equation will change)

$x + 7y - 8z = 38$

$5x - y + 2z = -2$ Add -5times eq. 1 to eq. 2

$2x - 3y \pm 7z = -21$ Add -2times eq. 1 to eq. 3

$x + 7y - 8z = 38$

$-36y + 42z = -192$

$-17y \pm 23z = -97$

The coefficients of y in the two equations are rather challenging for us. Dividing the second equation by -36 will produce unpleasant fractions, as will dividing the third equation by -17. Nevertheless, we proceed by choosing one of the alternatives, dividing the second equation by -36. Do not decimal approximate the results, as this will lead us to non-exact answers to the system.

$x + 7y - 8z = 38$

$y - (7/6)z = 48/9$

$-17y \pm 23z = -97$ Add 17 times the second equation to the third:

$x + 7y - 8z = 38$

$y - (7/6)z = 48/9$

$(19/6)z = -19/3$ Divide the terms of the third equation by 19/6

$x + 7y - 8z = 38$

$y - (7/6)z = 48/9$

$z = -2$

We perform back-substitution,

$$y - (7/6)(-2) = 48/9$$
$$y + 14/6 = 48/9$$
$$y = 48/9 - 14/6$$
$$y = 96/18 - 42/18$$
$$y = 54/18$$
$$y = 3$$
$$x + 7(3) - 8(-2) = 38$$
$$x + 21 + 16 = 38$$
$$x + 37 = 38$$
$$x = 1$$

The solution is the point $\{(1, 3, -2)\}$. We check these coordinates in the original system.

$$3x + 4y - z = 17$$
$$5x - y + 2z = -2$$
$$2x - 3y \pm 7z = -21$$
$$3(1) + 4(3) - (-2) = 17 \ \rightarrow \ 3 + 12 + 2 = 17$$
$$5(1) - 3 + 2(-2) = -2 \ \rightarrow \ 5 - 3 - 4 = -2$$
$$2(1) - 3(3) + 7(-2) = -21 \ \rightarrow \ 2 - 9 - 14 = -21$$

Note the coordinates of the point satisfy all the equations of the original system, and so we have found the solution.

Had any of the equations not been satisfied by the coordinates of the solution we found, then there is an error somewhere in our computations. Unfortunately, finding such a mistake would be a daunting task. Experience has taught your authors the best course of action in such a case would be to attempt the problem again from the start; this has the greatest likelihood of correcting a computational mistake.

Unusual things, mirroring the situation in systems of two linear equations, can happen to you in the process of doing Gaussian Elimination for three-variable systems of linear equations. For one thing, you can encounter the "No Solution," or "Inconsistent" case. This is detected in the same manner as was done for the two-equation systems. That is, at some point during the computation, a contradiction will appear.

$$x - 3y + z = 4$$
$$-x + 2y - 5z = 3$$
$$5x - 13y \pm 13z = 8$$

Since we have an x as the leading term of the first equation, we are set to begin. We will add 1 times the first equation to the second equation, and – 5 times the first equation to the third equation.

$$x - 3y + z = 4$$
$$-y - 4z = 72$$
$$y \pm 8z = -2$$

The second equation can have its leading coefficient changed to a 1 by multiplying the equation by −1:

$$x - 3y + z = 4$$
$$y + 4z = -72$$
$$y \pm 8z = -2$$

The y-term can be eliminated from the third equation by adding −2 times the second equation to the third:

$$x - 3y + z = 4$$
$$y + 4z = -7$$
$$0 = 12$$

Observe we have reached a contradiction. The last line of the system is now a contradiction, and this indicates to us the system has no solution.

The second unusual occurrence would be the appearance of the "Dependent" case, which is a special case of the consistent system. A dependent system has infinitely many solution points to it. The solutions to such a system are given "parametrically," and the somewhat unfortunate part about this is that more than one parametric representation for the solution to a dependent system may exist.

$$4x + 2y - 6z = 8$$
$$2x + y - 3z = 4$$
$$\_x - y \pm z = 4$$

Interchanging the first and third equations of the system yields:

$$x - y + z = 4$$
$$2x + y - 3z = 4$$
$$4x \pm 2y - 6z = 8$$

Now, add −2 times the first equation to the second, and −4 times the first equation to the third:

$$x - y + z = 4$$
$$3y - 5z = -4$$
$$6y - 10z = -8$$

Nothing is gained by interchanging the second and third equations, because division of those equations by their leading coefficient will yield fractions in either case. We leave the equations as they are, and divide the second equation by 3.

$$x - y + z = 4$$
$$y - (5/3)z = -4/3$$
$$\_6y - 10z = -8$$

Next, we add −6 times the second equation to the third:

$$x - y + z = 4$$
$$y - (5/3)z = -4/3$$
$$0 = 0$$

The final line is an identity. The system is dependent and has infinitely many solutions.

In the case of three-variable systems (and more than three variables!), the solution in the dependent case is expressed parametrically. We introduce a "dummy variable" t to the problem, and declare that $z = t$, where t can be any real number.

If this is the case, the second equation tells us $y - (5/3)t = -4/3$, or that $y = -4/3 + (5/3)t$.

Moving to the top equation of the system, we have $x - (-4/3 + (5/3)t) + t = 4$, or:

$$x + 4/3 - (2/3)\text{t} = 4$$
$$x = 8/3 + (2/3)\text{t}$$

The solutions to the system have the form

$(x, y, z) ? R^3 \mid x = 8/3 + 2/3t, y = -4/3 + 5/3t, z = t, fort ? R$

Let us demonstrate: if $t = 0$, the solution is (8/3, −4/3, 0). Substitution into the system equations yields:

$4x + 2y - 6z = 8$	$4(8/3) + 2(-4/3) - 0 = 8$	$32/3 - 8/3 = 8$	$24/3 = 8$
$2x + y - 3z = 4$	$2(8/3) + (-4/3) - 0 = 4$	$16/3 - 4/3 = 4$	$12/3 = 4$
$\_x - y \pm z = 4$	$8/3 - (-4/3) + 0 = 4$	$8/3 + 4/3 = 5$	$12/3 = 4$

For whatever value of t we insert into the parametric equations, we will always obtain a point in the solution of the system of three variables.

5.5 Introduction to Matrices

5.5.1 The Fundamentals of Matrices

Matrices are a fundamental tool in linear algebra and are used to represent and manipulate systems of linear equations. A matrix is a rectangular array of numbers, symbols, or expressions, arranged in horizontal rows and vertical columns. Matrices can be used to represent a wide variety of mathematical objects, including linear transformations, systems of linear equations, and more.

5.5.2 Definition and Terminology

A matrix is typically denoted by a capital letter, such as A, B, or C. The individual elements of a matrix are denoted by lowercase letters with subscripts, such as a_{ij}, where the subscripts i and j represent the row and column, respectively, in which the element lies. Should more than one matrix be present, we refer to them as "matrices," rather than "matrixes."

The size or dimension of a matrix is typically denoted as $m \times n$, where m is the number of rows and n is the number of columns. For example, a 3×4 matrix has 3 rows and 4 columns. Rows are numbered from top to bottom, columns from left to right.

In the particular case where a matrix happens to have exactly as many rows as it has columns, the matrix is referred to as a **square matrix**. The source of this term is the square shape taken on by the array when the number of rows and columns are equal.

Two matrices are said to be **equal** if they have precisely the same order, and entries in corresponding locations equal. In technical form, we can say $A = B$ if $ai, j = bi, j$ for all possible i,j entries.

There are some special matrices we should mention at this point. Any matrix (of whatever size) having all entries being 0 is referred to as a **zero matrix**. When you are interested in specifying a particular zero matrix, you might identify it (for instance) as the "3×4 zero matrix," possibly using the notation $0_{3,4}$, which would specify you were talking about a zero matrix of that specific size.

A square matrix having all entries being 0 except for 1 in the entries in the 1,1 position, the 2,2 position, the 3,3 position, and so on is referred to as an **identity matrix of order n**, where n is the number of rows in the matrix. Since the matrix is square, this also happens to be the number of columns in the matrix. An identity matrix is always identified as I_n, where n is the number of rows in the matrix.

The entries of a square matrix that are in the 1,1 position, the 2,2 position, the 3,3 position, and so on, form what we call the **main diagonal** entries of the square matrix. We will examine the main diagonal of a square matrix in greater detail shortly. Any matrix having all entries not lying on the main diagonal being 0, with main diagonal entries being either zero or nonzero (of whatever value) is called a **diagonal matrix**.

5.5.3 Matrix Operations

Matrices can be manipulated using algebraic operations. That is, they may be added, subtracted, multiplied, and scaled, just like numbers. Note the conspicuous absence of the operation of division – there is no division of matrices.

5.5.3.1 Matrix Addition and Subtraction

Addition and Subtraction: Two matrices can be added or subtracted if they have the same size (i.e., the same number of rows and columns). The resulting matrix will also have the same size, and the corresponding elements are added or subtracted.

The addition of two matrices is defined if and only if the matrices have the same size, that is, they have the same number of rows and the same number of columns. The sum of two matrices A and B, denoted $A + B$, is a matrix of the same size, where the corresponding elements are added.

Mathematically, if $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of size m × n, then their sum $A + B$ is a matrix of size m × n, where the (i, j)th element of $A + B$ is the sum of the corresponding elements in A and B, that is, $(A + B)_{ij} = a_{ij} + b_{ij}$.

Here is an example of matrix addition:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

$$C = A + B$$

To perform matrix addition, we add the corresponding elements of the matrices element-wise.

$$\begin{aligned} C &= A + B \\ &= \begin{bmatrix} 1+2 & 2+4 & 3+6 \\ 4+8 & 5+10 & 6+12 \\ 7+14 & 8+16 & 9+18 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix} \end{aligned}$$

Therefore, the resulting matrix C is:

$$C = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

In this example, the matrices A and B are both 3×3, so their sum C is also a 3×3 matrix, where the corresponding elements are added.

Similar to matrix addition, the subtraction of two matrices is defined if and only if the matrices have the same size. The difference of two matrices A and B, denoted $A - B$, is a matrix of the same size, where the corresponding elements are subtracted.

Mathematically, if $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of size m × n, then their difference $A - B$ is a matrix of size $m \times n$, where the (i, j)th element of $A - B$ is the difference of the corresponding elements in A and B, that is, $(A - B)_{ij} = a_{ij} - b_{ij}$.

Here is an example of matrix subtraction:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

$$C = A - B$$

To perform matrix subtraction, we subtract the corresponding elements of the matrices element-wise.

$$\begin{aligned} C &= A - B \\ &= [1-2 \quad 2-4 \quad 3-6] \\ &\quad [4-8 \quad 5-10 \quad 6-12] \\ &\quad [7-14 \quad 8-16 \quad 9-18] \\ &= [-1 \quad -2 \quad -3] \\ &\quad [-4 \quad -5 \quad -6] \\ &\quad [-7 \quad -8 \quad -9] \end{aligned}$$

Therefore, the resulting matrix C is:

$$\begin{aligned} C = &[-1 \quad -2 \quad -3] \\ &[-4 \quad -5 \quad -6] \\ &[-7 \quad -8 \quad -9] \end{aligned}$$

In this example, the matrices A and B are both 3×3, so their difference C is also a 3×3 matrix, where the corresponding elements are subtracted.

Matrix addition and subtraction have the following properties:

- Commutativity: $A + B = B + A$
- Associativity: $(A + B) + C = A + (B + C)$
- Existence of an additive identity: There exists a matrix 0 (the zero matrix) of the same size as A, such that $A + 0 = A$
- Existence of additive inverses: For every matrix A, there exists a matrix -A (the additive inverse of A) such that $A + (-A) = 0$

These properties are similar to the properties of addition and subtraction for real numbers and are crucial for understanding and manipulating systems of linear equations represented by matrices.

Example 5.23 Add the following matrices:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{bmatrix}$$

Solution:
To add two matrices, we add the corresponding elements in each position.

$$C = A + B$$

$$= \begin{bmatrix} 1+2 & 2+4 & 3+6 \\ 4+8 & 5+10 & 6+12 \\ 7+14 & 8+16 & 9+18 \end{bmatrix}$$

Simplifying, we get:

$$C = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

Therefore, the sum of the two matrices is:

$$C = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

Example 5.24 Subtract the following matrices:

$$A = \begin{bmatrix} 10 & 5 & 3 \\ 8 & 12 & 6 \\ 4 & 7 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 4 & 3 \\ 2 & 1 & 0 \end{bmatrix}$$

Solution:
To subtract two matrices, we subtract the corresponding elements in each position.

$$\begin{aligned} C &= A - B \\ &= \begin{bmatrix} 10-3 & 5-2 & 3-1 \\ 8-5 & 12-4 & 6-3 \\ 4-2 & 7-1 & 9-0 \end{bmatrix} \end{aligned}$$

Simplifying, we get:

$$C = \begin{bmatrix} 7 & 3 & 2 \\ 3 & 8 & 3 \\ 2 & 6 & 9 \end{bmatrix}$$

Therefore, the difference between the two matrices is:

$$C = \begin{bmatrix} 7 & 3 & 2 \\ 3 & 8 & 3 \\ 2 & 6 & 9 \end{bmatrix}$$

Example 5.25 Add the following matrices:

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 0 & -1 \\ -5 & 2 & 6 \end{bmatrix}$$

$$B = \begin{bmatrix} 2 & 1 & -1 \\ -3 & 4 & 2 \\ 1 & -2 & 3 \end{bmatrix}$$

Solution:
To add two matrices, we add the corresponding elements in each position.

$$\begin{aligned} C &= A + B \\ &= \begin{bmatrix} 1+2 & -2+1 & 3+(-1) \\ 4+(-3) & 0+4 & -1+2 \\ -5+1 & 2+(-2) & 6+3 \end{bmatrix} \end{aligned}$$

Simplifying, we get:

$$C = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 4 & 1 \\ -4 & 0 & 9 \end{bmatrix}$$

Therefore, the sum of the two matrices is:

$$C = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 4 & 1 \\ -4 & 0 & 9 \end{bmatrix}$$

5.5.3.2 Multiplication of a Matrix by a Scalar

A matrix can be multiplied by a scalar (a single number) by multiplying each element of the matrix by the scalar. The resulting matrix, denoted by kA, has the same size as the original matrix A.

Mathematically, if $A = [a_{ij}]$ is a matrix of size $m \times n$, and k is a scalar, then the matrix kA is also of size $m \times n$, where the (i, j)th element of kA is k times the (i, j)th element of A, that is, $(kA)_{ij} = k * a_{ij}$.

Here is an example of scalar multiplication of a matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$k = 3$$

$$B = 3A$$

To perform scalar multiplication, we multiply each element of the matrix A by the scalar $k = 3$.

$$B = 3A$$

$$= 3 * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

Therefore, the resulting matrix B is:

$$B = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

In this example, the scalar k is 3, and the resulting matrix B is obtained by multiplying each element of the original matrix A by 3.

Properties of Scalar Matrix Multiplication Scalar multiplication of matrices has the following properties:

- Distributivity with respect to matrix addition: $k(A + B) = kA + kB$
- Distributivity with respect to scalar addition: $(k + l)A = kA + lA$
- Associativity: $k(lA) = (kl)A$

These properties are similar to the properties of scalar multiplication for real numbers and are essential for manipulating and simplifying expressions involving matrices.

Applications of Scalar Matrix Multiplication Scalar multiplication of matrices has a wide range of applications, including:

- Image Processing: Scaling and transforming digital images by multiplying pixel values by a scalar.
- Numerical Analysis: Scaling the coefficients in systems of linear equations to improve the conditioning of the problem.
- Quantum Mechanics: Representing and transforming quantum states by multiplying state vectors by complex scalars.
- Computer Graphics: Scaling and rotating geometric objects by multiplying their coordinate matrices by scalars and transformation matrices.
- Economics: Scaling economic variables, such as prices and production levels, in input–output analysis and other economic models.

5.5.3.3 Matrix Multiplication

To this point, the operations with matrices have been somewhat intuitive, in the sense that the operations are performed in precisely the manner you would expect them to be done. Matrix multiplication is not quite so straightforward, though it is not particularly difficult to master.

Before introducing the method, we need to specify when the operation of matrix multiplication is defined, and the notation we will use. In general, matrix multiplication is denoted by juxtaposition of the matrices, or their names. That is, if one matrix is written directly alongside another, the implication is that the matrices are to be multiplied. Alternatively, if our matrices are named A and B, then if we write AB, we mean that we want to multiply matrix A with matrix B.

Two matrices can be multiplied if the number of columns in the first matrix is equal to the number of rows in the second matrix. The resulting matrix will have the same number of rows as the first matrix and the same number of columns as the second matrix. A convenient way to remember this product size is that if the first matrix A has size m × n and the second matrix B has size $n \times k$, the product matrix AB has size $m \times k$.

> Note: Matrix multiplication is not commutative. That is, using the sizes mentioned above, the product BA is not defined unless $n = m$. The product AB need not be the same size as matrix BA, and even if they are of the same size the product matrices are likely not going to be equal.

So, how does matrix multiplication proceed? When computing the i,j entry of the product matrix AB, we use row i of matrix A and column j of matrix B. The entries of row i are pairwise-multiplied with the entries of column j, one at a time, and the products are added together. The sum obtained will be entry i,j of the product matrix. That sounds mysterious, perhaps, and we will formalize it momentarily with what is called "double-subscripts" and "summation notation," but an example will make the procedure clear.

Here is an example of matrix multiplication:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$B = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

Find $C = A * B$.

Since matrix A has size 2×3 and matrix B has size 3×2, the product AB has size 2×2.

Entry $p_{1,1}$ is found by multiplying the entries of row 1 of matrix A against the entries of column 1 of matrix B, and adding those products. That would be $(1)(7) + (2)(9) + (3)(11)$. Note that we move across the row of matrix A and down the column of matrix B, in order. This simplifies to 58, and that is the 1,1 entry of the product matrix. The remaining entries of the product matrix are found in a similar manner, as illustrated below:

$C = [1 * 7 + 2 * 9 + 3 * 11 \; 1 * 8 + 2 * 10 + 3 * 12]$

$[4 * 7 + 5 * 9 + 6 * 11 \; 4 * 8 + 5 * 10 + 6 * 12]$

$C = [58 \quad 80]$

$[178 \quad 212]$

Therefore, the resulting matrix C is:

$C = [58 \quad 80]$

$[178 \quad 212]$

5.5.3.4 Formal Definition of Matrix Multiplication

The multiplication of two matrices is defined if the number of columns in the first matrix is equal to the number of rows in the second matrix. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the product AB is an $m \times p$ matrix.

Mathematically, the (i, j)th element of the product matrix AB is obtained by taking the dot product of the ith row of A and the jth column of B. Specifically, if $A = [a_{ij}]$ and $B = [b_{jk}]$, then the (i, j)th element of AB is:

$(AB)_{ij} = \Sigma\{k = 1\}\text{^}n \; a_{ik} * b_{kj}$

Here is an example of matrix multiplication:

$A = [1 \quad 2 \quad 3]$

$[4 \quad 5 \quad 6]$

$[7 \quad 8 \quad 9]$

$B = [1 \quad 4 \quad 7]$

$[2 \quad 5 \quad 8]$

$[3 \quad 6 \quad 9]$

C = AB

To perform matrix multiplication, we need to multiply each row of A by each column of B.

$C = AB$

$= [11 + 22 + 33 \quad 14 + 25 + 36 \quad 17 + 28 + 39]$

$[41 + 52 + 63 \quad 44 + 55 + 66 \quad 47 + 58 + 69]$

$[71 + 82 + 93 \quad 74 + 85 + 96 \quad 77 + 88 + 9*9]$

$= [14 \quad 32 \quad 50]$

$[32 \quad 77 \quad 122]$

$[50 \quad 122 \quad 194]$

Therefore, the resulting matrix C is:

$$C = \begin{bmatrix} 14 & 32 & 50 \\ 32 & 77 & 122 \\ 50 & 122 & 194 \end{bmatrix}$$

In this example, the matrix A has size 3×3, and the matrix B has size 3×3. The resulting matrix $C = AB$ has size 3×3, where each element is the dot product of the corresponding row of A and column of B.

Example 5.26 Multiply the following matrices:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$B = \begin{bmatrix} 5 & 1 \\ 2 & 3 \end{bmatrix}$$

Solution:
To multiply two matrices, the number of columns in the first matrix must be equal to the number of rows in the second matrix. The resulting matrix will have the same number of rows as the first matrix and the same number of columns as the second matrix.

The element in the ith row and jth column of the resulting matrix is obtained by taking the dot product of the ith row of the first matrix and the jth column of the second matrix.

$$C = A \times B$$

$$= \begin{bmatrix} (2 \times 5) + (3 \times 2) & (2 \times 1) + (3 \times 3) \\ (1 \times 5) + (4 \times 2) & (1 \times 1) + (4 \times 3) \end{bmatrix}$$

Simplifying, we get:

$$\mathrm{C} = \begin{bmatrix} 16 & 11 \\ 21 & 13 \end{bmatrix}$$

Therefore, the product of the two matrices is:

$$\mathrm{C} = \begin{bmatrix} 16 & 11 \\ 21 & 13 \end{bmatrix}$$

Example 5.27 Multiply the following matrices:

$$\mathrm{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$\mathrm{B} = \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}$$

Solution:
The number of columns in the first matrix (3) must be equal to the number of rows in the second matrix (3).

$$C = A \times B$$
$$= [(1 \times 7) + (2 \times 9) + (3 \times 11) \quad (1 \times 8) + (2 \times 10) + (3 \times 12)]$$
$$[(4 \times 7) + (5 \times 9) + (6 \times 11) \quad (4 \times 8) + (5 \times 10) + (6 \times 12)]$$

Simplifying, we get:

$$C = [58 \quad 70]$$
$$[167 \quad 202]$$

Therefore, the product of the two matrices is:

$$C = [58 \quad 70]$$
$$[167 \quad 202]$$

Example 5.28 Multiply the following matrices:

$$A = [1 \quad 2 \quad 3]$$
$$[4 \quad 5 \quad 6]$$
$$B = [7 \quad 8]$$
$$[9 \quad 10]$$
$$[11 \quad 12]$$

Solution:
The number of columns in the first matrix (3) must be equal to the number of rows in the second matrix (3).

$$C = A \times B$$
$$= [(1 \times 7) + (2 \times 9) + (3 \times 11) \; (1 \times 8) + (2 \times 10) + (3 \times 12)]$$
$$[(4 \times 7) + (5 \times 9) + (6 \times 11) \; (4 \times 8) + (5 \times 10) + (6 \times 12)]$$

Simplifying, we get:

$$C = [58 \quad 70]$$
$$[167 \quad 202]$$

Therefore, the product of the two matrices is:

$$C = [58 \quad 70]$$
$$[167 \quad 202]$$

These examples demonstrate the process of matrix multiplication. The key steps are: (1) ensure the number of columns in the first matrix is equal to the number of rows in the second matrix; (2) calculate the dot product of the *i*th row of the first matrix and the *j*th column of the second matrix to get the element in the *i*th row and *j*th column of the resulting matrix; and (3) simplify the resulting matrix.

Properties of Matrix Multiplication Matrix multiplication has the following important properties:

1) Associativity: $(AB)C = A(BC)$
2) Distributivity with respect to matrix addition: $A(B+C) = AB+AC$
3) Identity matrix: There exists a unique $n \times n$ matrix I, called the identity matrix, such that for any $n \times n$ matrix A, $AI = IA = A$.
4) Inverse matrix: If A is a square matrix and there exists a matrix B such that $AB = BA = I$, then B is called the inverse of A, denoted as A^{-1}.

These properties are crucial for manipulating and simplifying expressions involving matrix multiplication as well as for understanding the relationship between matrices and linear transformations.

Applications of Matrix Multiplication Matrix multiplication has a wide range of applications in various fields, including:

- Computer Graphics: Transforming and projecting 3D objects onto 2D screens by multiplying their coordinate matrices by transformation matrices.
- Quantum Mechanics: Representing and transforming quantum states by multiplying state vectors by unitary matrices.
- Cryptography: Encrypting and decrypting messages using matrix operations, as in the Hill cipher.
- Economics: Performing input–output analysis and modeling economic systems using matrix multiplication.
- Numerical Analysis: Solving systems of linear equations and computing the inverse of a matrix.
- Machine Learning: Performing operations such as linear regression, principal component analysis, and neural network computations using matrix multiplication.

5.5.3.5 The Multiplicative Inverse of a Square Matrix

In the previous section, we discussed the properties of matrix multiplication, including the concept of the identity matrix. In this section, we will explore the idea of the multiplicative inverse of a square matrix, which is a fundamental concept in linear algebra with numerous applications.

Definition of the Multiplicative Inverse A square matrix A is said to have a multiplicative inverse, denoted as $A\text{^}{-}1$, if there exists another square matrix B such that the product AB (or BA) is the identity matrix I. In other words, if A is an $n \times n$ matrix and there exists a matrix B of the same size such that:

$$\mathrm{AB} = \mathrm{BA} = \mathrm{I}$$

where I is the $n \times n$ identity matrix, then B is called the inverse of A, and A is called invertible.

Not all square matrices have a multiplicative inverse. Matrices that do have an inverse are called invertible or non-singular, while matrices that do not have an inverse are called singular.

Conditions for Invertibility For a square matrix A to be invertible, it must satisfy the following conditions:

- Determinant is nonzero: The determinant of the matrix A, denoted as $\det(A)$ or $|A|$, must be nonzero. If $det(A) = 0$, then the matrix is singular and does not have an inverse.

- Rank equals the number of rows/columns: The rank of the matrix A, which is the number of linearly independent rows or columns, must be equal to the number of rows (or columns) of the matrix. If the rank is less than the number of rows/columns, then the matrix is singular and does not have an inverse.

These conditions are necessary and sufficient for a square matrix to have a multiplicative inverse. If a matrix satisfies these conditions, then its inverse can be computed using various methods, as discussed in the next section.

Computing the Inverse of a Matrix There are several methods for computing the inverse of a matrix, including:

- Cofactor Method: This method involves computing the cofactors of the matrix elements and then forming the inverse matrix from these cofactors.
- Gaussian Elimination: This method involves performing row operations to transform the matrix into the identity matrix, with the intermediate steps yielding the inverse matrix.
- Inverse Matrix Formula: For a 2×2 matrix, the inverse can be computed using the formula:

$$\begin{aligned} &\mathrm{A} = [\mathrm{a} \quad \mathrm{b}] \\ &\quad\ [\mathrm{c} \quad \mathrm{d}] \\ &\mathrm{A}-1 = (1/[\mathrm{ad}-\mathrm{bc}]) * [\mathrm{d}- \quad \mathrm{b}] \\ &\qquad\qquad\qquad\quad [-\mathrm{c} \quad \mathrm{a}] \end{aligned}$$

- Inverse Matrix Theorem: If a matrix A is invertible, then its inverse can be computed as:

$$\mathrm{A}-1 = (1/\det(\mathrm{A}))^{*}\ \mathrm{adj}(\mathrm{A})$$

where adj(A) is the adjoint (or adjugate) of A, which is the transpose of the matrix of cofactors of A.

Example 5.29 Find the multiplicative inverse of the following matrix:

$$\begin{aligned} &\mathrm{A} = [2 \quad 3] \\ &\quad\ [1 \quad 4] \end{aligned}$$

Solution:
To find the multiplicative inverse of a square matrix A, we need to find a matrix B such that $A \times B = B \times A = I$, where I is the identity matrix of the same size as A.

Step 1: Calculate the determinant of matrix A.

$$\det(\mathrm{A}) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

Since the determinant is nonzero, the matrix A is invertible.

Step 2: Find the adjoint of matrix A.

The adjoint of a 2×2 matrix is obtained by interchanging the positions of the elements and changing the signs of the elements in the positions (1,2) and (2,1).

$$\begin{aligned} &\mathrm{Adj(A)} = [4 \quad -3] \\ &\qquad\quad\ [-1 \quad 2] \end{aligned}$$

Step 3: Calculate the multiplicative inverse of matrix A.

The multiplicative inverse of A is given by:

$$A - 1 = (1/\det(A)) \times \text{Adj}(A)$$

Substituting the values, we get:

$$A - 1 = (1/5) \times \begin{bmatrix} 4 & -3 \\ -1 & 2 \end{bmatrix}$$

Simplifying, we get:

$$A - 1 = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}$$

Therefore, the multiplicative inverse of matrix A is:

$$A - 1 = \begin{bmatrix} 4/5 & -3/5 \\ -1/5 & 2/5 \end{bmatrix}$$

Example 5.30 Find the multiplicative inverse of the following matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Solution:
Step 1: Calculate the determinant of matrix A.

$$\begin{aligned} \det(A) &= (1 \times 5 \times 9) + (2 \times 6 \times 7) + (3 \times 4 \times 8) - (3 \times 5 \times 7) - (1 \times 6 \times 8) - (2 \times 4 \times 9) \\ &= 45 + 84 + 96 - 105 - 48 - 72 \\ &= 0 \end{aligned}$$

Since the determinant is zero, the matrix A is not invertible.
Therefore, the multiplicative inverse of matrix A does not exist.

Example 5.31 Find the multiplicative inverse of the following matrix:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Solution:
Step 1: Calculate the determinant of matrix A.

$$\begin{aligned} \det(A) &= (2 \times 2 \times 2) - (2 \times (-1) \times (-1)) - ((-1) \times (-1) \times 2) - (0 \times 2 \times (-1)) \\ &\quad - (0 \times (-1) \times (-1)) - ((-1) \times 2 \times 0) \\ &= 8 - 4 - 2 - 0 - 0 - 0 \\ &= 2 \end{aligned}$$

Since the determinant is nonzero, the matrix A is invertible.

Step 2: Find the adjoint of matrix A.
The adjoint of a 3×3 matrix is obtained by taking the transpose of the cofactor matrix.

$$\begin{aligned}\text{Adj(A)} = &[2 \quad 1 \quad -1]\\ &[1 \quad 2 \quad -1]\\ &[-1 \quad -1 \quad 2]\end{aligned}$$

Step 3: Calculate the multiplicative inverse of matrix A.
The multiplicative inverse of A is given by:

$$\text{A} - 1 = (1/\det(\text{A})) \times \text{Adj(A)}$$

Substituting the values, we get:

$$\begin{aligned}\text{A} - 1 = (1/2) \times &[2 \quad 1 \quad -1]\\ &[1 \quad 2 \quad -1]\\ &[-1 \quad -1 \quad 2]\end{aligned}$$

Simplifying, we get:

$$\begin{aligned}\text{A} - 1 = &[1 \quad 1/2 \quad -1/2]\\ &[1/2 \quad 1 \quad -1/2]\\ &[-1/2 \quad -1/2 \quad 1]\end{aligned}$$

Therefore, the multiplicative inverse of matrix A is:

$$\begin{aligned}\text{A} - 1 = &[1 \quad 1/2 \quad -1/2]\\ &[1/2 \quad 1 \quad -1/2]\\ &[-1/2 \quad -1/2 \quad 1]\end{aligned}$$

These examples demonstrate the process of finding the multiplicative inverse of a square matrix. The key steps are: (1) calculate the determinant of the matrix; (2) find the adjoint of the matrix; and (3) calculate the multiplicative inverse using the formula $A^{-1} = (1/det(A)) \times Adj(A)$.

Applications of the Inverse Matrix The inverse matrix has many important applications in various fields, including:

- Solving Systems of Linear Equations: The inverse matrix can be used to solve systems of linear equations by multiplying both sides of the equation by the inverse of the coefficient matrix.
- Finding the Inverse of a Linear Transformation: The inverse matrix represents the inverse of a linear transformation, which is useful in fields like computer graphics, physics, and engineering.
- Computing Pseudo-inverses: In cases where a matrix is not invertible, the pseudo-inverse can be used to find the "best" approximate solution to a system of linear equations.
- Image Processing: Inverse matrices are used in image processing techniques, such as image deblurring and restoration.
- Cryptography: Inverse matrices are used in certain cryptographic algorithms, such as the Hill cipher, to encrypt and decrypt messages.
- Network Analysis: Inverse matrices are used in the analysis of electrical networks and the computation of network parameters.

Testing for Invertibility In the previous section, we discussed the concept of the multiplicative inverse of a square matrix and its various applications. In this section, we will explore the methods for determining whether a square matrix is invertible or not.

Determinant Test for Invertibility As mentioned earlier, a necessary and sufficient condition for a square matrix to be invertible is that its determinant must be nonzero. The determinant of a matrix, denoted as $\det(A)$ or $|A|$, is a scalar value that can be computed using various methods, such as the cofactor expansion or the Laplace expansion.

The determinant test for invertibility can be stated as follows:

- If $det(A) \neq 0$, then the matrix A is invertible.
- If $det(A) = 0$, then the matrix A is singular and not invertible.

Let us consider an example to illustrate the determinant test for invertibility.

Example 5.32 Determine whether the following matrix is invertible:

$$\begin{matrix} A = [2 & 3] \\ [4 & 5] \end{matrix}$$

To test for invertibility, we need to compute the determinant of the matrix A:

$$\det(A) = (2 \times 5) - (3 \times 4) = 10 - 12 = -2$$

Since the determinant of the matrix A is nonzero, the matrix is invertible.

Rank Test for Invertibility Another way to test for the invertibility of a square matrix is to use the rank test. Recall that the rank of a matrix is the number of linearly independent rows or columns in the matrix.

The rank test for invertibility can be stated as follows:

- If the rank of the matrix A is equal to the number of rows (or columns) of the matrix, then the matrix A is invertible.
- If the rank of the matrix A is less than the number of rows (or columns) of the matrix, then the matrix A is singular and not invertible.

The rank of a matrix can be computed using various methods, such as row echelon form or Gaussian elimination.

Example 5.33 Determine whether the following matrix is invertible:

$$\begin{matrix} B = [1 & 2 & 3] \\ [4 & 5 & 6] \\ [7 & 8 & 9] \end{matrix}$$

To determine if a matrix is invertible using the rank test, we follow these steps:

Find the rank of the matrix: The rank of a matrix is the maximum number of linearly independent rows (or columns) in the matrix. We can find the rank by performing row reduction (Gaussian elimination) on the matrix.

Compare the rank to the matrix size:

If the rank equals the number of rows (and columns, since it is a square matrix), the matrix is full rank and therefore invertible.

If the rank is less than the number of rows, the matrix is not full rank and therefore not invertible.

Let us apply this to matrix B:

B = [1 2 3]
[4 5 6]
[7 8 9]

Row Reduction:

Subtract multiples of the first row from the second and third rows to get zeros below the leading '1' in the first column:

[1 2 3]
[0 −3 −6] (R2 − 4R1)
[0 −6 −12] (R3 − 7R1)

Divide the second row by −3:

[1 2 3]
[0 1 2]
[0 −6 −12]

Add six times the second row to the third row:

[1 2 3]
[0 1 2]
[0 0 0]

Rank: We are left with two nonzero rows. Therefore, the rank of matrix B is 2.

The rank of B (2) is less than the number of rows in B (3). Therefore, matrix B is not invertible.

5.6.3 Inverse Matrix Theorem

The inverse matrix theorem provides a direct way to test for the invertibility of a square matrix:

A square matrix A is invertible if and only if $\det(A) \neq 0$.

If a matrix A is invertible, its inverse can be computed as:

$$A^{-1} = \frac{1}{\det(A)} \cdot \text{adj}(A)$$

where adj(A) is the adjoint (or adjugate) of the matrix A, which is the transpose of the matrix of cofactors of A.

Example 5.34 Compute the inverse of the following matrix:

C = [1 2]
[3 4]

To compute the inverse of the matrix C, we first need to calculate its determinant:

$$\det(C) = (1 \times 4) - (2 \times 3) = 4 - 6 = -2$$

Since the determinant is nonzero, the matrix C is invertible. We can now compute the inverse using the inverse matrix theorem:

C − 1 = (1/ det (C)) * adj(C)
= (1/ − 2) * [4 − 2]
[−3 1]
= [−1/2 1/2]
[3/2 − 1/2]

The inverse matrix $C^{\wedge}-1$ is the solution to the system of linear equations represented by the original matrix C.

Example 5.35 Determine if the following matrix is invertible:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

Solution:
To determine if a matrix is invertible, we need to check if its determinant is nonzero.
Step 1: Calculate the determinant of matrix A.

$$\begin{aligned} \det(A) &= (2 \times 4) - (3 \times 1) \\ &= 8 - 3 \\ &= 5 \end{aligned}$$

Since the determinant is nonzero, the matrix A is invertible.
Therefore, the matrix A is invertible.

Example 5.36 Determine if the following matrix is invertible:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Solution:
Step 1: Calculate the determinant of matrix A.

$$\begin{aligned} \det(A) &= (1 \times 5 \times 9) + (2 \times 6 \times 7) + (3 \times 4 \times 8) - (3 \times 5 \times 7) - (1 \times 6 \times 8) - (2 \times 4 \times 9) \\ &= 45 + 84 + 96 - 105 - 48 - 72 \\ &= 0 \end{aligned}$$

Since the determinant is zero, the matrix A is not invertible.
Therefore, the matrix A is not invertible.

Example 5.37 Determine if the following matrix is invertible:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Solution:
Step 1: Calculate the determinant of matrix A.

$$\begin{aligned} \det(A) &= (2 \times 2 \times 2) - (2 \times (-1) \times (-1)) - ((-1) \times (-1) \times 2) - (0 \times 2 \times (-1)) \\ &\quad - (0 \times (-1) \times (-1)) - ((-1) \times 2 \times 0) \\ &= 8 - 4 - 2 - 0 - 0 - 0 \\ &= 2 \end{aligned}$$

Since the determinant is nonzero, the matrix A is invertible.

Therefore, the matrix A is invertible.

The key step is to calculate the determinant of the matrix. If the determinant is nonzero, the matrix is invertible. If the determinant is zero, the matrix is not invertible.

For a square matrix to be invertible, its determinant must be nonzero. If the determinant is zero, the matrix is not invertible and does not have a multiplicative inverse.

5.5.4 Using Matrices to Solve Systems of Linear Equations

We introduced the concept of matrices and explored various properties related to matrix invertibility. In this chapter, we will focus on using matrices to solve systems of linear equations.

5.5.4.1 Representing a System of Linear Equations as a Matrix Equation

A system of linear equations can be represented in matrix form as $Ax = b$, where A is the coefficient matrix, x is the vector of unknowns, and b is the vector of constants.

Consider the following system of linear equations:

$$2x + 3y = 8$$

$$4x + 5y = 14$$

We can represent this system in matrix form as:

$$[2 \quad 3]\,[x] = [8]$$

$$[4 \quad 5]\,[y] = [14]$$

or, in compact form:

$$Ax = b$$

where:

$$A = [2 \quad 3]$$
$$\quad\;\; [4 \quad 5]$$
$$x = [x]$$
$$\quad\;\; [y]$$
$$b = [8]$$
$$\quad\;\; [14]$$

The matrix A is the coefficient matrix, x is the vector of unknowns, and b is the vector of constants.

Example 5.38 Solving a 2×2 System of Linear Equations

Let us solve the system of linear equations represented by the matrix equation $Ax = b$:

$$[2 \quad 3]\,[x] = [8]$$

$$[4 \quad 5]\,[y] = [14]$$

To solve this system, we can use the inverse of the coefficient matrix A:

$$x = A - 1\ ^{*}\ b$$

First, let us compute the inverse of the matrix A:

$$\det(A) = (2 * 5) - (3 * 4) = 10 - 12 = -2$$

$$A - 1 = (1/\det(A)) * \text{adj}(A)$$

$$= (1/-2) * [5 - 3]$$
$$[-4 \quad 2]$$

$$= [-5/2 \quad 3/2]$$
$$[2 \quad -1]$$

Now, we can compute the solution vector x:

$$x = A - 1 * b$$

$$= [-5/2 \quad 3/2] * [8]$$
$$[2 \quad -1] \; [14]$$

$$= [-1]$$
$$[3]$$

Therefore, the solution to the system of linear equations is $x = -1$ and $y = 3$.

Example 5.39 Solving a 3×3 System of Linear Equations

Consider the following system of linear equations:

$$x + 2y + 3z = 10$$
$$2x + 3y + 4z = 14$$
$$3x + 4y + 5z = 18$$

We can represent this system in matrix form as:

$$[1 \quad 2 \quad 3] \; [x] = [10]$$
$$[2 \quad 3 \quad 4] \; [y] = [14]$$
$$[3 \quad 4 \quad 5] \; [z] = [18]$$

or, in compact form:

$$Ax = b$$

where:

$$A = [1 \quad 2 \quad 3]$$
$$[2 \quad 3 \quad 4]$$
$$[3 \quad 4 \quad 5]$$

$$x = [x]$$
$$[y]$$
$$[z]$$

$$b = [10]$$
$$[14]$$
$$[18]$$

To solve this system, we can use the inverse of the coefficient matrix A:

$$x = A - 1 * b$$

First, let us compute the inverse of the matrix A:

$$\det(A) = 1(15 - 12) - 2(10 - 15) + 3(8 - 6) = 3 - 10 + 6 = -1$$

$$A - 1 = (1/\det(A)) * \text{adj}(A)$$

$$= (1/-1) * [15 \quad -12 \quad 8]$$
$$[-10 \quad 15 \quad -6]$$
$$[8 \quad -10 \quad 6]$$

$$= [-15 \quad 12 \quad -8]$$
$$[10 \quad -15 \quad 6]$$
$$[-8 \quad 10 \quad -6]$$

Now, we can compute the solution vector x:

$$x = A - 1 * b$$

$$= [-15 \quad 12 \quad -8] * [10]$$
$$[10 \quad -15 \quad 6] \; [14]$$
$$[-8 \quad 10 \quad -6] \; [18]$$

$$= [2]$$
$$[3]$$
$$[1]$$

Therefore, the solution to the system of linear equations is $x = 2$, $y = 3$, and $z = 1$.

5.5.4.2 Augmented Matrix and Gaussian Elimination

Another way to solve a system of linear equations is by using the augmented matrix and Gaussian elimination.

The augmented matrix of a system of linear equations is the matrix formed by appending the constant vector b to the coefficient matrix A. For the system $Ax = b$, the augmented matrix is $[A|b]$.

For the 2×2 system of linear equations example:

$$[2 \quad 3 \mid 8]$$
$$[4 \quad 5 \mid 14]$$

The augmented matrix represents the entire system in a compact form.

Gaussian elimination is a method for solving systems of linear equations by transforming the augmented matrix into row echelon form. The steps involved in Gaussian elimination are:

1) Convert the leading entry of each row to 1 by dividing the row by the leading entry.
2) Use the leading 1 in each row to eliminate the entries below it in the same column.
3) Repeat steps 1 and 2 until the augmented matrix is in row echelon form.

Once the augmented matrix is in row echelon form, the system of linear equations can be solved by back-substitution.

Example 5.40 Solving a 2×2 System of Linear Equations Using Gaussian Elimination

Let us solve the 2×2 system of linear equations using Gaussian elimination:

$$2x + 3y = 8$$

$$4x + 5y = 14$$

The augmented matrix for this system is:

$$[2 \quad 3 \mid 8]$$

$$[4 \quad 5 \mid 14]$$

Step 1: Convert the leading entry of the first row to 1 by dividing the first row by 2.

$$[1 \quad 3/2 \mid 4]$$

$$[4 \quad 5 \quad \mid 14]$$

Step 2: Use the leading 1 in the first row to eliminate the entry in the first row, second column of the second row.

$$[1 \quad 3/2 \mid 4]$$

$$[0 \quad 1/2 \mid 6]$$

Step 3: Convert the leading entry of the second row to 1 by dividing the second row by 1/2.

$$[1 \quad 3/2 \quad \mid 4]$$

$$[0 \quad 1 \quad \mid 12]$$

The augmented matrix is now in row echelon form. We can solve the system by back-substitution:

$$x = 4$$

$$y = 12$$

Therefore, the solution to the system of linear equations is $x = 4$ and $y = 12$.

Example 5.41 Solving a 3×3 System of Linear Equations Using Gaussian Elimination

Consider the following 3×3 system of linear equations:

$$x + 2y + 3z = 10$$

$$2x + 3y + 4z = 14$$

$$3x + 4y + 5z = 18$$

The augmented matrix for this system is:

$$[1 \quad 2 \quad 3 \mid 10]$$

$$[2 \quad 3 \quad 4 \mid 14]$$

$$[3 \quad 4 \quad 5 \mid 18]$$

Next, we can perform row operations to transform the matrix into row echelon form, which will allow us to solve for the variables x, y, and z.

First, let us multiply the first row by -2 and add it to the second row to eliminate the x term in the second equation:

$$[1 \quad 2 \quad 3 \quad | \ 10]$$
$$[0 \ -1 \ -2 \ | -6]$$
$$[3 \quad 4 \quad 5 \quad | \ 18]$$

Now, let us multiply the first row by -3 and add it to the third row to eliminate the x term in the third equation:

$$[1 \quad 2 \quad 3 \quad | \ 10]$$
$$[0 \ -1 \ -2 \ | -6]$$
$$[0 \ -2 \ -4 \ | -12]$$

Next, we will multiply the second row by 2 to get a leading 1 in the second equation:

$$[1 \quad 2 \quad 3 \quad | \ 10]$$
$$[0 \ -2 \ -4 \ | -12]$$
$$[0 \ -2 \ -4 \ | -12]$$

Finally, we will divide the second row by -2 to get a leading 1 in the second equation:

$$[1 \quad 2 \quad 3 \quad | \ 10]$$
$$[0 \quad 1 \quad 2 \quad | \ 6]$$
$$[0 \ -2 \ -4 \ | -12]$$

Now we can solve for the variables. From the second row, we can see that $y = 6$.

Substituting $y = 6$ into the first equation, we get:

$$x + 2(6) + 3z = 10$$
$$x + 12 + 3z = 10$$
$$x + 3z = -2$$

Substituting $y = 6$ into the third equation, we get:

$$3x + 4(6) + 5z = 18$$
$$3x + 24 + 5z = 18$$
$$3x + 5z = -6$$

Now we have a 2×2 system of equations in x and z:

$$x + 3z = -2$$
$$3x + 5z = -6$$

Solving this system, we get:

$$x = 2$$
$$z = -2$$

Therefore, the solution to the original 3×3 system of linear equations is:

$$x = 2$$
$$y = 6$$
$$z = -2$$

Example 5.42 Solving a 3×3 System with Unique Solution
Consider the following system of linear equations:

$$2x + 3y - z = 7$$
$$-x + y + 2z = 5$$
$$3x - 2y + z = 1$$

We can represent this system in matrix form as:

$$[2 \quad 3 \quad -1 \mid 7]$$
$$[-1 \quad 1 \quad 2 \mid 5]$$
$$[3 \quad -2 \quad 1 \mid 1]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 0 \quad 0 \mid 2]$$
$$[0 \quad 1 \quad 0 \mid 3]$$
$$[0 \quad 0 \quad 1 \mid 1]$$

The solution is:

$$x = 2$$
$$y = 3$$
$$z = 1$$

This system has a unique solution.

Example 5.43 Solving a 3×3 System with Infinitely Many Solutions
Consider the following system of linear equations:

$$x + 2y - z = 3$$
$$2x + 4y - 2z = 6$$
$$3x + 6y - 3z = 9$$

We can represent this system in matrix form as:

$$[1 \quad 2 \quad -1 \mid 3]$$
$$[2 \quad 4 \quad -2 \mid 6]$$
$$[3 \quad 6 \quad -3 \mid 9]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 2 \quad -1 \mid 3]$$
$$[0 \quad 0 \quad 0 \mid 0]$$
$$[0 \quad 0 \quad 0 \mid 0]$$

The solution is:

$$x = t$$
$$y = -t/2$$
$$z = t$$

where t is any real number. This system has infinitely many solutions.

Example 5.44 Solving a 3×3 System with No Solution
Consider the following system of linear equations:

$$x + y + z = 2$$
$$2x + 2y + 2z = 4$$
$$3x + 3y + 3z = 6$$

We can represent this system in matrix form as:

$$[1 \quad 1 \quad 1 \mid 2]$$
$$[2 \quad 2 \quad 2 \mid 4]$$
$$[3 \quad 3 \quad 3 \mid 6]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 1 \quad 1 \mid 2]$$
$$[0 \quad 0 \quad 0 \mid 0]$$
$$[0 \quad 0 \quad 0 \mid 2]$$

The last row of the augmented matrix shows that the constant term is not equal to 0, which means the system has no solution.

Example 5.45 Solving a 4×4 System with Unique Solution
Consider the following system of linear equations:

$$2x + 3y - z + 4w = 20$$
$$-x + y + 2z - w = 7$$
$$4x - 2y + 3z + 2w = 16$$
$$3x + y - z - w = 5$$

We can represent this system in matrix form as:

$$[2 \quad 3 \quad -1 \quad 4 \mid 20]$$
$$[-1 \quad 1 \quad 2 \quad -1 \mid 7]$$
$$[4 \quad -2 \quad 3 \quad 2 \mid 16]$$
$$[3 \quad 1 \quad -1 \quad -1 \mid 5]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 0 \quad 0 \quad 0 \mid 2]$$
$$[0 \quad 1 \quad 0 \quad 0 \mid 3]$$
$$[0 \quad 0 \quad 1 \quad 0 \mid 1]$$
$$[0 \quad 0 \quad 0 \quad 1 \mid 4]$$

The solution is:

$$x = 2$$
$$y = 3$$
$$z = 1$$
$$w = 4$$

This system has a unique solution.

Example 5.46 Solving a 3×3 System with No Solution

Consider the following system of linear equations:

$$2x + 3y - z = 7$$
$$-x + y + 2z = 5$$
$$3x - 2y + z = 1$$

We can represent this system in matrix form as:

$$[2 \quad 3 - 1 \mid 7]$$
$$[-1 \quad 1 \quad 2 \mid 5]$$
$$[3 - 2 \quad 1 \mid 1]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 0 \quad 0 \mid 2]$$
$$[0 \quad 1 \quad 0 \mid 3]$$
$$[0 \quad 0 \quad 1 \mid 4]$$

The last row of the augmented matrix shows that the constant term is not equal to 0, which means the system has no solution.

Example 5.47 Solving a 3×3 System with Infinitely Many Solutions

Consider the following system of linear equations:

$$x + y + z = 3$$
$$2x + 2y + 2z = 6$$
$$3x + 3y + 3z = 9$$

We can represent this system in matrix form as:

$$[1 \quad 1 \quad 1 \mid 3]$$
$$[2 \quad 2 \quad 2 \mid 6]$$
$$[3 \quad 3 \quad 3 \mid 9]$$

Performing Gaussian elimination, we get the augmented matrix in row echelon form:

$$[1 \quad 1 \quad 1 \mid 3]$$
$$[0 \quad 0 \quad 0 \mid 0]$$
$$[0 \quad 0 \quad 0 \mid 0]$$

The solution is:

$$x = t$$
$$y = -t$$
$$z = 2\,t$$

where t is any real number. This system has infinitely many solutions.

Example 5.48 Solve the following system of linear equations using matrices:

$$2x + 3y = 8$$

$$x + 4y = 7$$

Solution:

Step 1: Write the system of linear equations in matrix form.

Let us define the coefficient matrix A, the constant vector b, and the variable vector x as follows:

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

$$b = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$$

$$x = \begin{bmatrix} x \\ y \end{bmatrix}$$

The system of linear equations can be written in matrix form as:

$$Ax = b$$

Step 2: Find the inverse of the coefficient matrix A.

To solve the system of linear equations, we need to find the inverse of the coefficient matrix A.

$$\det(A) = (2 \times 4) - (3 \times 1) = 8 - 3 = 5$$

Since the determinant is nonzero, the matrix A is invertible.

The inverse of matrix A is:

$$A^{-1} = \begin{bmatrix} 4/5 & -3/5 \\ 1/5 & 2/5 \end{bmatrix}$$

Step 3: Multiply both sides of the matrix equation by the inverse of A.

$$Ax = b$$

$$A^{-1} \times Ax = A^{-1} \times b$$

$$x = A^{-1} \times b$$

Substituting the values, we get:

$$x = \begin{bmatrix} 4/5 * 8 + (-3/5) * 7 & 1/5 * 8 + 2/5 * 7 \end{bmatrix}$$

$$= \begin{bmatrix} 4 - 6/5 & 2/5 + 14/5 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Therefore, the solution to the system of linear equations is:

$$x = 2$$

$$y = 4$$

Example 5.49 Solve the following system of linear equations using matrices:

$$3x + 2y - z = 5$$
$$2x - y + 3z = 11$$
$$x + y - z = 3$$

Solution:

Step 1: Write the system of linear equations in matrix form.

Let us define the coefficient matrix A, the constant vector b, and the variable vector x as follows:

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 2 & -1 & 3 \\ 1 & 1 & -1 \end{bmatrix}$$
$$b = \begin{bmatrix} 5 \\ 11 \\ 3 \end{bmatrix}$$
$$x = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The system of linear equations can be written in matrix form as:

$$Ax = b$$

Step 2: Find the inverse of the coefficient matrix A.

To solve the system of linear equations, we need to find the inverse of the coefficient matrix A.

$$\begin{aligned} \det(A) &= (3 \times (-1) \times (-1)) + (2 \times 1 \times 3) + ((-1) \times (-1) \times 1) - (3 \times (-1) \times 1) \\ &\quad - (2 \times 1 \times (-1)) - (1 \times 1 \times 3) \\ &= 3 + 6 + 1 + 3 + 2 + 3 \\ &= 18 \end{aligned}$$

Since the determinant is nonzero, the matrix A is invertible.

The inverse of matrix A is:

$$A - 1 = \begin{bmatrix} 1/3 & 1/6 & -1/6 \\ 1/3 & -1/3 & 1/3 \\ 1/6 & 1/6 & -1/6 \end{bmatrix}$$

Step 3: Multiply both sides of the matrix equation by the inverse of A.

$$Ax = b$$
$$A - 1 \times Ax = A - 1 \times b$$
$$x = A - 1 \times b$$

Substituting the values, we get:

$$x = \begin{bmatrix} 1/3 & 1/6 & -1/6 \\ 1/3 & -1/3 & 1/3 \\ 1/6 & 1/6 & -1/6 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Therefore, the solution to the system of linear equations is:

$$x = 1$$
$$y = 2$$
$$z = 1$$

Example 5.50 Consider the following system of linear equations:

$$2x + 3y = 12$$
$$5x - y = 7$$

To solve this system using matrices, we can represent the system as:

$$\begin{bmatrix} 2 & 3 \\ 5 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

Step 1: Create the coefficient matrix A and the constant vector b.

$$A = \begin{bmatrix} 2 & 3 \\ 5 & -1 \end{bmatrix}$$
$$b = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

Step 2: Use Gaussian elimination to convert the coefficient matrix A into row echelon form.

$$\begin{bmatrix} 2 & 3 \\ 5 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3/2 \\ 0 & -11/2 \end{bmatrix}$$

Step 3: Use backward substitution to solve for the variables x and y.
From the second row, we have: $-11/2y = 7$
Solving for y, we get: $y = -2/11$
Substituting y into the first row, we have:

$$2x + 3(-2/11) = 12$$
$$2x - 6/11 = 12$$
$$2x = 12 + 6/11$$
$$x = 6 + 3/11$$

Therefore, the solution to the system is:

$$x = 6 + 3/11$$
$$y = -2/11$$

Example 5.51 Consider the following system of linear equations:

$$2x + 3y - z = 8$$
$$x - y + 2z = 5$$
$$3x + 2y + z = 12$$

To solve this using matrices, we can represent the system as:

$$\begin{bmatrix} 2 & 3 & -1 \\ 1 & -1 & 2 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 12 \end{bmatrix}$$

Step 1: Create the coefficient matrix A and the constant vector b.

$$\begin{aligned}
\text{A} &= [2 \quad 3 \quad -1] \\
&\quad [1 \quad -1 \quad 2] \\
&\quad [3 \quad 2 \quad 1] \\
\text{b} &= [8] \\
&\quad [5] \\
&\quad [12]
\end{aligned}$$

Step 2: Use Gaussian elimination to convert the coefficient matrix A into row echelon form.

$$\begin{aligned}
&[2 \quad 3 \quad -1] \quad\quad [1 \quad 3/2 \quad -1/2] \\
&[1 \quad -1 \quad 2] = [0 \quad -5/2 \quad 3] \\
&[3 \quad 2 \quad 1] \quad\quad [0 \quad 0 \quad 1]
\end{aligned}$$

Step 3: Use backward substitution to solve for the variables x, y, and z.
From the third row, we have: $z = 12 - 2y$
Substituting this into the second row, we get:

$$\begin{aligned}
&-5/2y + 3(12 - 2y) = 5 \\
&-5/2y + 36 - 6y = 5 \\
&-11/2y = -31 \\
&y = 2
\end{aligned}$$

Substituting $y = 2$ into the third row, we get:

$$z = 12 - 2(2) = 8$$

Substituting $y = 2$ and $z = 8$ into the first row, we get:

$$\begin{aligned}
&2x + 3(2) - 8 = 8 \\
&2x + 6 - 8 = 8 \\
&2x = 10 \\
&x = 5
\end{aligned}$$

Therefore, the solution to the system is:

$$\begin{aligned}
&x = 5 \\
&y = 2 \\
&z = 8
\end{aligned}$$

These examples demonstrate how to use matrices to solve systems of linear equations. The key steps are:

1) Represent the system as a matrix equation $Ax = b$.
2) Use Gaussian elimination to convert the coefficient matrix A into row echelon form.
3) Use backward substitution to solve for the variables.

6

Sequences and Series

A sequence is an ordered list of numbers, where each number in the list is called a term of the sequence. The order of the terms of the sequence cannot be changed, and those terms may or may not follow some sort of an observable or describable pattern. Sequences can be finite, where the number of terms is limited, or infinite, where the number of terms goes on forever.

The general form of a sequence is:

$$a_1, a_2, a_3, \ldots, a_n, \ldots$$

where a_n represents the nth term of the sequence. The subscript n is called the index of the sequence, and it starts from 1 and goes up to infinity for an infinite sequence. There are times when, for the sake of description using a formula, the sequence may begin with term zero, represented as a_0. An example of this will be presented shortly.

When describing or referring to a sequence whose nth term is given by an, the notation used is either $\{a_n\}$ or $\{a_n\}_{n=1}^{n=k}$. The value of k is infinity if the sequence is infinite, and it is a constant if the sequence is finite.

6.1 Sequences

There are different ways to define a sequence. The most common are:

- Explicit Formula: Each term of the sequence is defined by a formula that depends on the index n as the independent variable. For example, the sequence 2, 4, 6, 8, … can be defined by the formula $a_n = 2n$. The formula to produce the terms of a sequence may not be unique; that is, there are occasions when different formulas produce the same ordered list of numbers.
- Recursive Formula: Each term of the sequence is defined in terms of the previous terms. For example, the sequence 1, 3, 5, 7, … can be defined by the recursive formula $a_n = a_{n-1} + 2$, where $a_1 = 1$.
- Tables or Graphs: A sequence can be defined by listing the values of the terms or by plotting the terms as distinct points on a graph, where the x-axis is used to represent the index values of the terms of the sequence.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

6.1.1 Sequences Given by Explicit Formulas Using the Index as the Independent Variable

Example 6.1 Find the first five terms of the sequence defined by the formula $a_n = 3n-1$.

Solution:

- $a_1 = 3(1)-1 = 2$
- $a_2 = 3(2)-1 = 5$
- $a_3 = 3(3)-1 = 8$
- $a_4 = 3(4)-1 = 11$
- $a_5 = 3(5)-1 = 14$

The first five terms of the sequence are: 2, 5, 8, 11, 14.

Example 6.2 Find the first five terms of the sequence given by the formula $a_n = (-1)^{n+2}$.

Solution:

$$a_1 = (-1)^3 = -1$$
$$a_2 = (-1)^4 = 1$$
$$a_3 = (-1)^5 = -1$$
$$a_4 = (-1)^6 = 1$$
$$a_1 = (-1)^7 = -1$$

The first five terms of the sequence are: $-1, 1, -1, 1, -1$.

Example 6.3 Find the first five terms of the sequence given by the formula $a_n = \cos n\pi$.

Solution:

$$a_1 = \cos \pi = -1$$
$$a_2 = \cos 2\pi = 1$$
$$a_3 = \cos 3\pi = -1$$
$$a_4 = \cos 4\pi = 1$$
$$a_1 = \cos 5\pi = -1$$

The first five terms of the sequence are: $-1, 1, -1, 1, -1$.

Observe, in the previous two examples given, the terms of the sequence are identical, yet the formulas producing the terms of the sequences are entirely different. It is important to remember one can devise multiple "generating formulas" for the terms of a sequence, some possibly more elegant than others, and some computationally more useful than others.

6.1.2 Sequences Given by a Recursive Formula

Recursion is a situation where events "fall back upon previous events." That is, what happens at any given point is determined by what happened before.

A recursive formula is a sequence-generating formula that produces terms based upon one or more of its predecessors. To define a sequence using a recursive formula, knowledge must be available regarding the values of earlier terms in the sequence.

Example 6.4 Find the first five terms of the sequence defined by the recursive formula $a_n = a_{n-1} + 3$, where $a_1 = 1$.

Solution:

- $a_1 = 1$
- $a_2 = a_1 + 3 = 1 + 3 = 4$
- $a_3 = a_2 + 3 = 4 + 3 = 7$
- $a_4 = a_3 + 3 = 7 + 3 = 10$
- $a_5 = a_4 + 3 = 10 + 3 = 13$

The first five terms of the sequence are: 1, 4, 7, 10, 13.

Example 6.5 Find the first five terms of the sequence defined by the recursive formula $a_n = a_{n-1} - 3$, given $a_1 = 7$.

Solution:

- $a_1 = 7$
- $a_2 = a_1 - 3 = 7 - 3 = 4$
- $a_3 = a_2 - 3 = 4 - 3 = 1$
- $a_4 = a_3 - 3 = 1 - 3 = -2$
- $a_5 = a_4 - 3 = -2 - 3 = -5$

The first five terms of the sequence are: 7, 4, 1, −2, −5.

6.1.3 Comparison of Generating Formulas and Recursive Formulas

The drawback of a recursive formula lies in the need for terms to be calculated using predecessors. If, for instance, we took the example $a_n = a_{n-1} + 3$, where $a_1 = 1$, and asked for the 1000th term of the sequence, it would be necessary to first calculate the initial 999 terms of the sequence. A generating formula does not require knowledge of the preceding terms; thus, the 1000th term of the sequence given by $a_n = 3n - 1$ is $a_{1000} = 3(1000) - 1 = 2999$.

It is occasionally possible to determine a generating formula from a sequence given recursively, though this is not always a simple task (particularly when first attempting to do so). Much trial and error is involved, and a fair amount of ingenuity. This is compounded by the fact the generating formula is not necessarily unique.

For the example of the sequence defined by the recursive formula $a_n = a_{n-1} - 3$, given $a_1 = 7$, we can see the terms of sequence begin at 7 and decrease successively by 3. This is a constant decrease in value from term to term, which suggests a linear rate of decrease. A possible generating formula can be found by determining the slope–intercept form of the equation of the line passing through the points (1, 7), (2, 4), (3, 1), (4, −2), and (5, −5).

Recalling our work on the equations of lines, it can be seen a possible example of the generating formula is $a_n = 10 - 3n$.

6.1.4 Sequences Given by Tables

A sequence presented in this manner resembles a table of values (as used when graphing equations), where the independent variable represents the index. Unlike a table of values, we do not infer from the table a continuous graph in the plane. If and when we plot such a sequence, the graph will consist of discrete points in the plane.

> *Note: Beware the appearance of homonyms*! *"Discrete" points in the plane are not "discreet" points in the plane*!

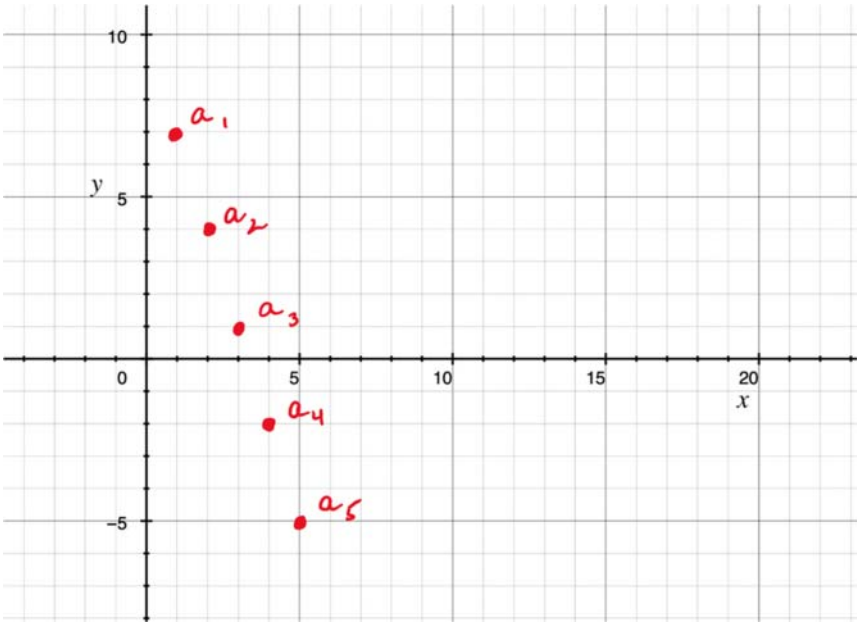
Example 6.6 Shown are the first five terms of an infinite sequence given in a table.

n	a_n
1	7
2	4
3	1
4	−2
5	−5

Observe this is the same sequence we saw given by the recursive formula $a_n = a_{n-1} - 3$, given $a_1 = 7$, and by the generating formula $a_n = 10 - 3n$. It is not obvious the three representations refer to the same sequence, and you should not take it to be so.

6.1.5 Sequences Given by Graphs of Discrete Points in the Plane

We can plot the terms of a sequence as discrete points in the plane having coordinates (n, a_n). For the example shown as an illustration of the terms of a sequence given by a table, we obtain the graph:



There are times when the visual representation of a sequence can be very informative. Each of the representations has its own purpose and value, and thus it is important to recognize the different types that exist.

6.2 Arithmetic Sequences

An arithmetic sequence is a sequence where the difference between consecutive terms is constant. This constant difference is called the common difference, and it is denoted by the letter d. The general form of the terms of an arithmetic sequence is:

$$a_1, a_1 + d, a_1 + 2d, \ldots, a_1 + (n-1)d, \ldots$$

where

- a_1 is the first term of the sequence.
- d is the common difference.

This can be more succinctly expressed by the explicit Formula for an arithmetic sequence.

$$a_n = a_1 + w\,(n-1)d$$

where

- a_n is the nth term of the sequence.
- a_1 is the first term of the sequence.
- d is the common difference.

Example 6.7 Find the first five terms of the arithmetic sequence with $a_1 = 3$ and $d = 2$.

Solution: The explicit formula for the arithmetic sequences is $a_n = 3 + (n-1)(2)$. Using this formula, we can find the first five terms:

$$a_1 = 3 + (1-1)(2) = 3 + 0 = 3$$
$$a_2 = a_1 + (2-1)(2) = 3 + 2 = 5$$
$$a_3 = a_1 + (3-1)(2) = 3 + 2(2) = 7$$
$$a_4 = a_1 + (4-1)(2) = 3 + 3(2) = 9$$
$$a_5 = a_1 + (5-1)(2) = 3 + 4(2) = 11$$

Therefore, the first five terms of the sequence are: 3, 5, 7, 9, 11.

The nth term of an arithmetic sequence can be found using the generating formula, just as we did earlier. The reason for this is that the particular generating formula of an arithmetic sequence is nothing more than a special instance of the sort of generating formula we saw before.

Example 6.8 Find the 10th term of the arithmetic sequence with first term $a_1 = 3$ and common difference $d = 2$.

To find the 10th term, a_{10}, we can use the explicit formula:

$$a_{10} = a_1 + (n-1)d$$
$$a_{10} = 3 + (10-1)(2)$$

$$a_{10} = 3 + 18$$
$$a_{10} = 21$$

Therefore, the 10th term of the arithmetic sequence is 21.

Example 6.9 Find the 206th term of the arithmetic sequence with first term $a_1 = -4$ and common difference $d = -3$.

To find the 20th term, a_{20}, we can use the explicit formula:

$$a_{206} = a_1 + (n-1)d$$
$$a_{206} = -4 + (206-1)(-3)$$
$$a_{206} = -4 + (205)(3)$$
$$a_{206} = -4 + 615$$
$$a_{206} = 611$$

Therefore, the 206th term of the arithmetic sequence is 611.

Example 6.10 Finding the 50th term of an arithmetic sequence.

Suppose we have an arithmetic sequence with first term $a_1 = 5$ and common difference $d = 4$.

To find the 50th term, a_{50}, we can use the explicit formula:

$$a_{50} = a_1 + (n-1)d$$
$$a_{50} = 5 + (50-1)4$$
$$a_{50} = 5 + 196$$
$$a_{50} = 201$$

Therefore, the 50th term of the arithmetic sequence is 201.

A novel application of algebra related to sequences is the problem of finding the term number for a particular value within a sequence.

Let us say we have an arithmetic sequence with the first term $a_1 = 2$ and the common difference $d = 3$. We want to find the index number for which the arithmetic sequence attains a particular value, such as 35. Because we know the structure of the generating formula for an arithmetic sequence, we can use the explicit formula for the terms of the sequence to determine the point at which the desired term appears.

Recall the explicit formula for an arithmetic sequence:

$$a_n = a_1 + (n-1)d$$

We are interested in the value n that produces an output of 35.

$$35 = 2 + (n-1)(3)$$
$$35 = 2 + 3n - 3$$
$$33 = 3n$$
$$n = 11$$

The value 35 appears as the 11th term of the arithmetic sequence.

Example 6.11 For the sequence whose nth term is given by $a_n = \left(\frac{2}{7}\right)^{n-1}$, find the value of n that produces the term having the value $\frac{256}{5{,}764{,}801}$.

Solution: The required value of n will satisfy the equation

$$\left(\frac{2}{7}\right)^{n-1} = \frac{256}{5{,}764{,}801}$$

At first, this may seem a daunting task, due to the immensity of the denominator. However, the power rule of exponents tells us

$$\left(\frac{2}{7}\right)^{n-1} = \frac{2^{n-1}}{7^{n-1}}$$

The value of n must satisfy both of the following equations:

$$2^{n-1} = 256$$

and

$$7^{n-1} = 5{,}764{,}801$$

It is our choice which of these to solve, and then all that remains is to verify that the solution also satisfies the other equation. It appears the first equation will be easier to solve, so we will attempt to do so.

$$2^{n-1} = 256$$

Exponential equations such as this can be solved through conversion to logarithmic form or through an appropriate rewrite of the non-exponential value in terms of the base for the exponential value. We will perform both techniques to serve as a review; each, in fact, utilizes the same observation.

$$n - 1 = \log_2 256$$

We observe $256 = 2^8$, and so

$$n - 1 = \log_2 2^8$$

The definition of logarithms tells us this simplifies to

$$n - 1 = 8$$

or

$$n = 9$$

We confirm $7^{9-1} = 7^8 = 5{,}764{,}801$ and thus have found $\frac{256}{5{,}764{,}801}$ to be the ninth term of the sequence.

The same result is found by detecting $256 = 2^8$ at the outset and thus appealing to the principle of equality as it relates to exponents of the same base:

$$2^{n-1} = 256$$
$$2^{n-1} = 2^8$$
$$n - 1 = 8$$
$$n = 9$$

6.2.1 Summation Notation

Summation notation, also known as sigma notation, is a way to represent the sum of a sequence of terms. The general form of a summation is:

$$\sum_{i=a}^{b} a_i$$

where

- $\sum$ is the summation symbol (this is the capital letter sigma in the Greek alphabet).
- i is the index variable that represents the terms being summed.
- a is the lower bound of the index (the starting value of i).
- b is the upper bound of the index (the ending value of i).
- a_i is the general term of the sequence being summed.

Summation notation is a shorthand expression describing the result of the addition of a progression of numbers whose values are determined by the general term of the sequence being summed.

Example 6.12 Evaluate the sum $\sum_{i=1}^{15}(2i+1)$

Solution: To evaluate the sum, we need to substitute the values of i from the lower bound (1) to the upper bound (15) and then add the resulting terms.

$$\begin{aligned}\sum_{i=1}^{15}(2i+1) &= (2\times 1+1)+(2\times 2+1)+(2\times 3+1)+\cdots+(2\times 15+1)\\ &= 3+5+7+9+11+13+15+17+19+21+23+25+27+29+31\\ &= 225\end{aligned}$$

The key steps are:

- Identify the index of summation (i) and its range (1 – 15).
- Substitute the values of i into the expression ($2i+1$).
- Add up all the resulting terms.

This summation represents the sum of an arithmetic sequence with a common difference of 2, starting from 3 and ending at 31.

Naturally, the procedure shown in the example is impractical if the difference between the lower and upper bounds of summation is substantial. For instance, had we been asked to find $\sum_{i=1}^{1000}(2i+1)$, the task would have involved a tedious but straightforward computation. Fortunately, there exist techniques that allow us to avoid such inconvenience, which we will discuss shortly.

6.2.1.1 Finding the Sum of Terms of a Finite Series

A series is called finite if the upper bound of the summation is a finite value. While there is no general method for finding the sum of terms of a finite series, there are some frequently used series, which have been studied in sufficient detail that formulas exist for determining the sum of the first n terms of that series. We present these formulas without proof and will use them as accepted truths. Here, k represents a constant, and n is some natural number greater than 1.

$$\sum_{i=1}^{n} k = kn$$

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=1}^{n} i^3 = \frac{n^2(n+1)^2}{4}$$

$$\sum_{i=1}^{n} (a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$$

$$\sum_{i=1}^{n} ca_i = c\sum_{i=1}^{n} a_i$$

Example 6.13 Find

$$\sum_{i=1}^{260} 3i$$

Solution:

$$\sum_{i=1}^{260} 3i = 3\sum_{i=1}^{260} i = 3\left[\frac{(260)(261)}{2}\right] = 101{,}790$$

Example 6.14 Find

$$\sum_{i=1}^{90} (2i^2 - 5i + 7)$$

Solution:

$$\sum_{i=1}^{90} (2i^2 - 5i + 7)$$

$$\sum_{i=1}^{90} (2i^2) - \sum_{i=1}^{90} (5i) + \sum_{i=1}^{90} (7)$$

$$2\sum_{i=1}^{90} (i^2) - 5\sum_{i=1}^{90} (i) + \sum_{i=1}^{90} (7)$$

$$2\left[\frac{(90)(91)(181)}{6}\right] - 5\left[\frac{(90)(91)}{2}\right] + 7(90)$$

$$494{,}130 - 20{,}475 + 630$$

$$474{,}285$$

6.2.1.2 Sum of the First *n* Terms of an Arithmetic Sequence

The sum of the first n terms of an arithmetic sequence can be found using the formula:

$$S_n = (n/2)[2a_1 + (n-1)d]$$

where

- S_n is the sum of the first n terms.
- n is the number of terms.
- a_1 is the first term.
- d is the common difference.

Example 6.15 Find the sum of the first 10 terms of the arithmetic sequence with $a_1 = 2$ and $d = 3$.

Solution: Using the formula for the sum of the first n terms, we can find the sum:

$$S_n = (n/2)[2a_1 + (n-1)d]$$
$$S_{10} = (10/2)[2(2) + (10-1)(3)]$$
$$S_{10} = 5[4 + 27]$$
$$S_{10} = 5(31)$$
$$S_{10} = 155$$

Therefore, the sum of the first 10 terms of the sequence is 155.

Example 6.16 Finding the sum of the first 10 terms of an arithmetic sequence with the first term $a_1 = 3$ and the common difference $d = 2$.

Solution: We want to find the sum of the first 10 terms of this sequence.

The explicit formula for the nth term of an arithmetic sequence is:

$$a_n = a_1 + (n-1)d$$

To find the sum of the first 10 terms, we can use the formula:

$$S_n = n/2 * (a_1 + a_n)$$

Substituting the values, we get:

$$S_{10} = 10/2 * (3 + (10-1)2)$$
$$S_{10} = 5 * (3 + 18)$$
$$S_{10} = 5 * 21$$
$$S_{10} = 105$$

Therefore, the sum of the first 10 terms of the arithmetic sequence with $a_1 = 3$ and $d = 2$ is 105.

Example 6.17 Finding the sum of the first 20 terms of an arithmetic sequence with $a_1 = -2$ and $d = 3$.

Solution: Let us apply the formula to find the sum of the first 20 terms of an arithmetic sequence with $a_1 = -2$ and $d = 3$.

Using the formula:

$$S_{20} = 20/2 * (2(-2) + (20-1)3)$$
$$S_{20} = 10 * (-4 + 57)$$
$$S_{20} = 10 * 53$$
$$S_{20} = 530$$

Therefore, the sum of the first 20 terms of the arithmetic sequence with $a_1 = -2$ and $d = 3$ is 530.

Example 6.18 Finding the sum of the first 50 terms of an arithmetic sequence with the first term $a_1 = 7$ and the common difference $d = 4$.

Solution: Using the formula:

$$S_n = n/2 * (a_1 + a_n)$$

First, we need to find the 50th term, a_{50}, using the explicit formula:

$$a_{50} = a_1 + (n-1)d$$
$$a_{50} = 7 + (50-1)4$$
$$a_{50} = 7 + 196$$
$$a_{50} = 203$$

Now, we can plug in the values into the sum formula:

$$S_{50} = 50/2 * (7 + 203)$$
$$S_{50} = 25 * 210$$
$$S_{50} = 5250$$

Therefore, the sum of the first 50 terms of the arithmetic sequence with $a_1 = 7$ and $d = 4$ is 5250.

Example 6.19 Finding the sum of the first n terms of an arithmetic sequence with negative terms with the first term $a_1 = -5$ and the common difference $d = -3$.

Solution: Using the formula:

$$S_n = n/2 * (a_1 + a_n)$$

First, let us find the nth term, a_n, using the explicit formula:

$$a_n = a_1 + (n-1)d$$
$$a_n = -5 + (n-1)(-3)$$
$$a_n = -5 - 3n + 3$$
$$a_n = -3n - 2$$

Now, we can plug in the values into the sum formula:

$$S_n = n/2 * (-5 - 3n - 2)$$
$$S_n = n/2 * (-5n - 2)$$
$$S_n = (-5n^2)/2 - n$$

This is the formula for the sum of the first n terms of the arithmetic sequence with $a_1 = -5$ and $d = -3$.

6.2.1.3 Finding the Sum of the First *n* Terms of an Arithmetic Sequence with Varying Common Differences

Let us consider an arithmetic sequence where the common difference changes after the first few terms.

For example, let us say the sequence starts with $a_1 = 2$, $d_1 = 3$ for the first five terms and then changes to $d_2 = 5$ for the remaining terms.

The sequence is defined as follows:
First five terms:

$$a_1 = 2$$
$$d_1 = 3$$

Remaining terms:

$$d_2 = 5$$

To find the sum of the first n terms, we can use the formula for the sum of an arithmetic sequence:

$$\text{Sum} = n/2 * (2a_1 + (n-1)d)$$

where

n is the number of terms.
a_1 is the first term.
d is the common difference.

For the first five terms:

$$n = 5$$
$$a_1 = 2$$
$$d = d_1 = 3$$

Plugging these values into the formula:

$$\text{Sum} = 5/2 * (2 * 2 + (5-1) * 3)$$
$$\text{Sum} = 5/2 * (4 + 12)$$
$$\text{Sum} = 5/2 * 16$$
$$\text{Sum} = 40$$

For the remaining terms ($n - 5$ terms):

$$n = n - 5$$
$$a_1 = a_6 = 2 + 5 * 3 = 17$$
$$d = d_2 = 5$$

Plugging these values into the formula:

$$\text{Sum} = (n-5)/2 * (2 * 17 + (n-5-1) * 5)$$
$$\text{Sum} = (n-5)/2 * (34 + 5(n-6))$$
$$\text{Sum} = (n-5)/2 * (34 + 5n - 30)$$
$$\text{Sum} = (n-5)/2 * (4 + 5n)$$

The total sum of the first n terms is the sum of the first five terms plus the sum of the remaining $(n - 5)$ terms:

$$\text{Total sum} = 40 + (n-5)/2 * (4 + 5n).$$

6.3 Geometric Sequences

A geometric sequence is a sequence where each term is obtained by multiplying the previous term by a common ratio. Formally, a geometric sequence is a sequence a_n, where each term after the first is obtained by multiplying the previous term by a constant, non-zero number r, called the common ratio.

The general formula for a geometric sequence is:

$$a_n = a_1 * r^{(n-1)}$$

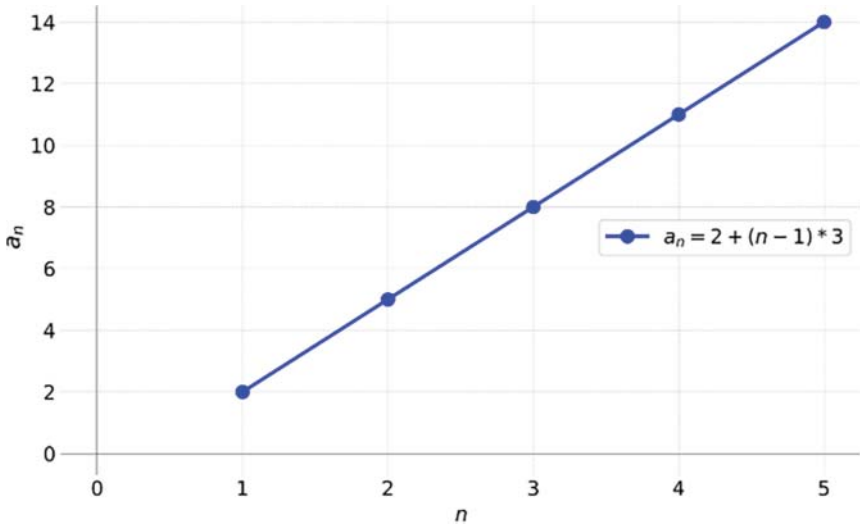
where

- a_n is the nth term of the sequence.
- a_1 is the first term of the sequence.
- r is the common ratio.

For example, consider the sequence: 2, 6, 18, 54, 162, ...

Here, the first term $a_1 = 2$, and the common ratio $r = 3$, since each term is three times the previous term. We can write this sequence using the general formula:

$$a_n = 2 * 3^{(n-1)}$$



The graph above shows the first six terms of this geometric sequence. Notice how each term is three times the previous term, demonstrating the constant common ratio.

6.3.1 Infinite Geometric Sequences

An infinite geometric sequence is one where the number of terms goes on indefinitely. The sum of the terms in an infinite geometric sequence is given by the formula:

$$S = a_1/(1 - r)$$

where

- S is the sum of the infinite sequence.
- a_1 is the first term.
- r is the common ratio.

For the example sequence above, if we want to find the sum of all the terms, we can use the formula:

$$S = 2/(1 - 3) = 2/(-2) = -1$$

So, the sum of the infinite sequence 2, 6, 18, 54, 162, ... is −1.

Many real-world situations can be modeled by geometric series. One illustration is the growth of a population over a finite span of time. It is important to note the growth of a population cannot perpetually experience exponential growth. The reason for this is the limitation of resources, which restricts the growth of that population (this concept is integral to the notion of the "carrying capacity" of the environment). However, over a relatively short span of time, growth can be approximated by a geometric model.

Example 6.20 The population of a town is growing at a rate of 5% per year. If the initial population is 10,000 people, find the population after 10 years and the sum of the population over the next 20 years.

Solution: Let us define the variables:

- Initial population: $a_1 = 10{,}000$.
- Common ratio: $r = 1.05$ (since the population grows by 5% per year).
- Number of years: $n = 10$.

Step 1: Find the population after 10 years.

Using the formula for the nth term of a geometric sequence:

$$a_n = a_1 * r^{(n-1)}$$

$$a_{10} = 10{,}000 * 1.05^{(10-1)} = 10{,}000 * 1.05^9 = 16{,}288.55$$

Step 2: Find the sum of the population over the next 20 years.

The sum of an infinite geometric series is given by the formula:

$$S = a_1/(1 - r)$$

where

- S is the sum of the series.
- a_1 is the first term.
- r is the common ratio.

Substituting the values:

$$S = 10{,}000/(1 - 1.05) = 10{,}000/(-0.05) = 200{,}000$$

Therefore, the population after 10 years is 16,288.55 people, and the sum of the population over the next 20 years is 200,000 people.

A second application frequently encountered is the impact of inflation and deflation on prices. This notion is significant because the intuitive view of price increases and reductions is easily seen to be inaccurate.

Consider a store having a weekend sale over the span of which prices are reduced by 50% per day for 3 days. The intuitive analysis of this situation seems bizarre because two consecutive 50% reductions in price would seem (at first glance) to reduce the price of goods to zero. Clearly, this cannot be

the case, and thus our intuitive understanding of successive price changes must be fundamentally flawed.

Example 6.21 A company is offering a promotion where the price of a product is discounted by 15% each month. If the initial price of the product is \$100, find the price of the product after 6 months and the sum of the discounts over the next 12 months.

Solution:

1) Identify the geometric sequence.
   First term (a): The initial price of the product is \$100.
   Common ratio (r): The price is discounted by 15%, meaning each month the price is 85% of the previous month's price. So, the common ratio is 0.85.
2) Price after 6 months.
   The price each month forms a geometric sequence: \$100, \$85, \$72.25, ...
   To find the price after 6 months (the 6th term in the sequence), we can use the formula:

$$n\text{th term} = \text{a} * r^{(n-1)}$$

   where
   a is the first term,
   r is the common ratio,
   n is the term number.
   Plugging in the values:

$$6\text{th term} = \$100 * (0.85)^{(6-1)}$$

$$6\text{th term} = \$100 * (0.85)^5$$

$$6\text{th term} \approx \$37.71$$

   Therefore, the price of the product after 6 months is approximately \$37.71.3.
3) Sum of discounts over 12 months.
   Price after 12 months: First, calculate the price of the product after 12 months using the same formula:

$$12\text{th term} = \$100 * (0.85)^{(12-1)}$$

$$12\text{th term} = \$100 * (0.85)^{11}$$

$$12\text{th term} \approx \$14.23$$

   Total discount: Subtract the final price (after 12 months) from the initial price:

$$\text{Total discount} = \$100 - \$14.23 = \$85.77$$

   Therefore, the sum of the discounts over 12 months is approximately \$85.77.

6.3.2 Convergence of Infinite Geometric Sequences

For an infinite geometric sequence to converge (i.e. have a finite sum), the common ratio $|r|$ must be less than 1. If $|r| > 1$, then the sequence diverges and the sum is infinite.

For example, the sequence 2, 6, 18, 54, ... has a common ratio of 3, which is greater than 1, so it diverges and has an infinite sum.

However, the sequence 100, 50, 25, 12.5, ... has a common ratio of 1/2, which is less than 1, so it converges and has a finite sum of 200.

6.4 Fibonacci Sequence

The Fibonacci sequence is a famous sequence of numbers that has wide-ranging applications in mathematics, science, art, and nature. This sequence was first described by the Italian mathematician Leonardo of Pisa, also known as Fibonacci, in the 13th century. The Fibonacci sequence is defined as follows:

The first two terms of the sequence are 0 and 1, and each subsequent term is the sum of the two preceding terms. Mathematically, the Fibonacci sequence can be expressed as:

$$F(n) = F(n-1) + F(n-2)$$

where $F(0) = 0$ and $F(1) = 1$.

The first few terms of the Fibonacci sequence are:

0,1,1,2,3,5,8,13,21,34,55,89,144, …

As you can see, the sequence starts with 0 and 1, and each subsequent number is the sum of the two preceding numbers.

6.4.1 Properties of the Fibonacci Sequence

The Fibonacci sequence exhibits several interesting properties:

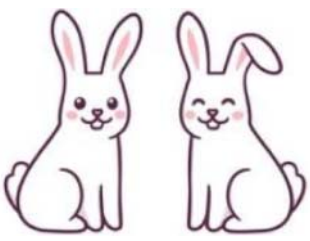
1) Golden ratio: The ratio of consecutive Fibonacci numbers approaches the golden ratio, ϕ (phi), as the numbers get larger. The golden ratio is approximately 1.618033988749895.
2) Recursive formula: The Fibonacci sequence can be expressed recursively as $F(n) = F(n-1) + F(n-2)$, with $F(0) = 0$ and $F(1) = 1$.
3) Closed-form formula: The Fibonacci sequence can also be expressed using a closed-form formula known as the Binet formula: $F(n) = (\phi^n - (1-\phi)^n)/\sqrt{5}$.
4) Divisibility: The Fibonacci numbers have interesting divisibility properties. For example, any Fibonacci number is divisible by a Fibonacci number with a lower index if and only if the index of the larger number is a multiple of the index of the smaller number.
5) Appearance in nature: The Fibonacci sequence can be observed in various natural phenomena, such as the arrangement of leaves on a stem, the pattern of seeds in a sunflower, and the spiral shape of seashells.

The essence of the sequence is captured in the analysis of the growth of a population of creatures, such as bunnies, under particular assumptions.

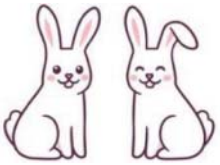
Let us suppose we consider the number of pairs of adult bunnies, assuming that there is some breeding period for bunnies that is the same for any particular pair of bunnies—for simplicity, let us call it a month. We will start with one pair, and assume it takes 1 month for bunnies to mature to adulthood.

For the first month, we have one pair of adult bunnies, and at the end of that month, we have one pair of adult bunnies and one pair of baby bunnies.

Month 1:



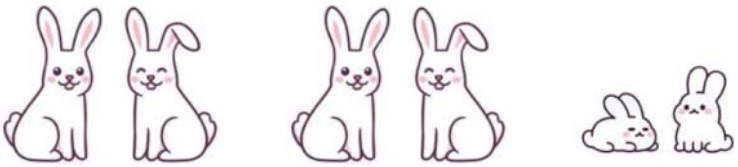
We have one pair of adult bunnies.
Start of Month 2:



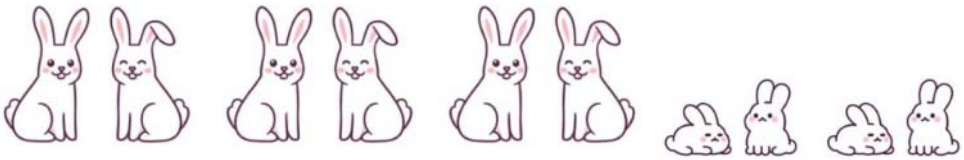
Note, we still have one pair of adult bunnies.



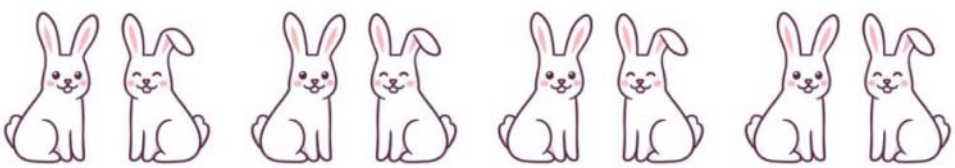
By the end of month 2, the adult bunnies are still adults, have had another pair of baby bunnies, and the first pair of babies are now adults. We now have this situation:

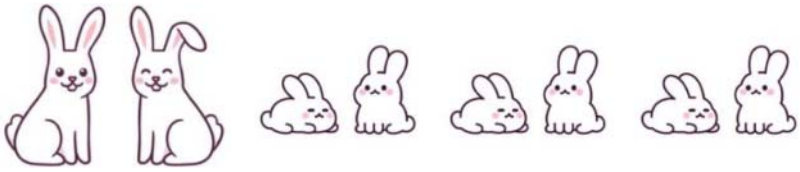


Observe we now have two pairs of adult bunnies, and one pair of baby bunnies.
Third month: Each of the adult pairs of bunnies has babies, and the pair of baby bunnies above grows to adulthood. We now have:



Observe there are now THREE pairs of adult bunnies.
Fourth month: The three adult pairs each have babies, and the two sets of baby bunnies grow to adulthood. Now we have:





See, we now have five pairs of adult bunnies.

In the next month, all five pairs of adults have babies, and the three sets of baby bunnies grow to be adults. That means we have:





There are now eight pairs of adult bunnies and five pairs of baby bunnies.

Look back now at the growth of the number of adult pairs, listing the number of pairs in each month:

Month	**1**	**2**	**3**	**4**	**5**	**6**
# of adult bunny pairs	**1**	**1**	**2**	**3**	**5**	**8**

Observe the number of adult bunny pairs per month is precisely the Fibonacci sequence.

Example 6.22 Find the 10th term of the Fibonacci sequence.

Solution: The Fibonacci sequence is a series of numbers in which each number is the sum of the two preceding ones, usually starting with 0 and 1. The sequence is as follows:

$$0,1,1,2,3,5,8,13,21,34,\ldots$$

To find the 10th term, we can use the following formula:

$$F(n) = F(n-1) + F(n-2)$$

where

$F(n)$ is the nth term of the Fibonacci sequence
$F(n-1)$ is the $(n-1)$th term.
$F(n-2)$ is the $(n-2)$th term.

Plugging in the values, we get:

$$\begin{aligned} F(10) &= F(9) + F(8) \\ &= 34 + 21 \\ &= 55 \end{aligned}$$

Therefore, the 10th term of the Fibonacci sequence is 55.

Example 6.23 Prove that the sum of the first n Fibonacci numbers is equal to the $(n + 2)$th Fibonacci number minus 1.

Solution: Let us denote the Fibonacci numbers as $F(n)$, where n is the position in the sequence.
We can express the sum of the first n Fibonacci numbers as:

$$S(n) = F(1) + F(2) + F(3) + \cdots + F(n)$$

To prove the given statement, we can use the following steps:
Step 1: Express $S(n)$ in terms of Fibonacci numbers.

$$\begin{aligned} S(n) &= F(1) + F(2) + F(3) + \cdots + F(n) \\ &= 1 + 1 + 2 + 3 + 5 + \cdots + F(n) \end{aligned}$$

Step 2: Recognize the pattern in the Fibonacci sequence.

$$F(n + 2) = F(n + 1) + F(n)$$

Step 3: Rearrange the terms to get the desired expression.

$$S(n) = F(n + 2) - 1$$

Therefore, the sum of the first n Fibonacci numbers is equal to the $(n + 2)$th Fibonacci number minus 1.

6.4.2 Applications of the Fibonacci Sequence

The Fibonacci sequence has a wide range of applications, including:

1) Computer science: The Fibonacci sequence is used in various computer science algorithms, such as the Fibonacci heap, which is a priority queue data structure.
2) Finance: The Fibonacci sequence is used in technical analysis of financial markets, where traders often look for Fibonacci retracement levels to identify potential support and resistance levels.
3) Art and design: The golden ratio, which is closely related to the Fibonacci sequence, is often used in art, architecture, and design to create aesthetically pleasing compositions.
4) Biology: The Fibonacci sequence can be observed in the growth patterns of various organisms, such as the arrangement of leaves on a stem, the spiral pattern of seeds in a sunflower, and the spiral shape of seashells.
5) Mathematics: The Fibonacci sequence is used in the study of number theory, combinatorics, and other areas of mathematics.

6.4.3 Applications of Fibonacci Sequence in Cybersecurity

The Fibonacci sequence, named after the medieval Italian mathematician Leonardo of Pisa, also known as Fibonacci, is a captivating mathematical pattern that has found numerous applications in various fields, including cybersecurity. This sequence, where each number is the sum of the two preceding ones, starting with 0 and 1, has intriguing properties that make it particularly useful in the realm of information security.

6.4.3.1 Cryptography and Encryption

One of the key applications of the Fibonacci sequence in cybersecurity is its use in cryptography and encryption algorithms. The predictable yet complex nature of the Fibonacci sequence can be leveraged to create strong encryption ciphers. For example, the Fibonacci-based encryption algorithm, also known as the Fibonacci cipher, uses the Fibonacci sequence to generate a key stream that is then combined with the plaintext using an XOR operation, providing a robust and secure method of data encryption.

The Fibonacci cipher is particularly effective because the Fibonacci sequence is not easily predictable, making it challenging for attackers to break the encryption. Additionally, the vast number of possible Fibonacci-based encryption keys, driven by the exponential growth of the sequence, further enhances the security of this approach.

6.4.3.2 Password Generation and Management

Another application of the Fibonacci sequence in cybersecurity is in the generation and management of strong passwords. The inherent complexity and unpredictability of the Fibonacci sequence can be used to create unique, complex, and hard-to-guess passwords. By incorporating Fibonacci-based patterns into password generation algorithms, security professionals can ensure that the resulting passwords are both memorable and highly secure.

The Fibonacci sequence can be used to implement password policies that require users to change their passwords at intervals based on the Fibonacci numbers. This approach helps to maintain the overall security of the system by regularly updating the passwords while also making the process more intuitive for users.

6.4.3.3 Network Security and Traffic Analysis

The Fibonacci sequence can also be leveraged in the realm of network security and traffic analysis. Researchers have explored the use of Fibonacci-based patterns to detect anomalies in network traffic, identify potential cyber threats, and optimize network resource allocation.

One example is the application of Fibonacci-based algorithms in network intrusion detection systems (NIDS). By analyzing the patterns of network traffic and identifying deviations from expected Fibonacci-based patterns, NIDS can effectively detect and alert on potential security breaches and unauthorized activities.

The Fibonacci sequence can be used to model the growth and distribution of network resources, such as IP addresses and bandwidth allocation, to ensure efficient and secure resource management.

6.5 The Principles of Mathematical Induction

The principle of mathematical induction is a powerful tool in the study of sequences and series, allowing us to prove statements about an infinite set of numbers or objects by considering a base

case and then showing that a general step holds. This method of proof is particularly useful when dealing with recursive definitions, where each term in a sequence is defined in terms of the previous terms.

The Principle of Mathematical Induction

The principle of mathematical induction can be stated as follows:

To prove that a statement P(n) is true for all positive integers n, it is sufficient to:

1) Prove that the statement P(1) is true (the base case).
2) Prove that if the statement P(k) is true for some positive integer k, then the statement P($k + 1$) is also true (the inductive step).

The logic behind this principle is that if the base case is true and the inductive step holds, then the statement P(n) must be true for all positive integers n.

Example 6.24 Proving the formula for the sum of the first n positive integers.

Let's consider the statement: "The sum of the first n positive integers is given by the formula S(n) = n(n+1)/2."

To prove this statement using mathematical induction, we follow the two steps outlined above:

Base case ($n = 1$):

When $n = 1$, the formula gives S(1) = 1(1 + 1)/2 = 1, which is the correct sum of the first positive integer.

Inductive step:

Assume that the statement is true for some positive integer k, that is, S(k) = k($k + 1$)/2.
We need to show that the statement is also true for $k + 1$, that is, S($k + 1$) = ($k + 1$)(($k + 1$) + 1)/2.

$$\begin{aligned} S(k+1) &= S(k) + (k+1) \\ &= k(k+1)/2 + (k+1) \\ &= (k^2 + 3\,k + 2)/2 \\ &= (k+1)(k+2)/2 \\ &= (k+1)((k+1)+1)/2 \end{aligned}$$

Therefore, the statement is true for $k + 1$, and by the principle of mathematical induction, the formula for the sum of the first n positive integers is true for all positive integers n.

Example 6.25 Prove that the sum of the first n positive integers is equal to $n(n + 1)/2$.

Solution: To prove this statement using the Principle of Mathematical Induction, we will follow these steps:

Step 1: Base case

Let us start by verifying the statement for the first positive integer, $n = 1$.
The sum of the first 1 positive integer is 1.
The expression $n(n + 1)/2$ for $n = 1$ is $1(1 + 1)/2 = 1$.
The base case is true.

Step 2: Inductive step

Assume that the statement is true for some positive integer $n = k$.
That is, the sum of the first k positive integers is $k(k + 1)/2$.
Now, we need to prove that the statement is also true for $n = k + 1$.

The sum of the first $(k+1)$ positive integers is the sum of the first k positive integers plus $(k+1)$.
Sum of the first $(k+1)$ positive integers = Sum of the first k positive integers + $(k+1)$

$$= k(k+1)/2 + (k+1)$$
$$= (k^2 + 3\,k + 2)/2$$
$$= (k+1)(k+2)/2$$

This matches the expression $n(n+1)/2$ for $n = k+1$.
Therefore, the statement is true for all positive integers n.

Example 6.26 Prove that $2^n > n$ for all positive integers n.

Solution: To prove this statement using the Principle of Mathematical Induction, we will follow these steps:

Step 1: Base case
Let us start by verifying the statement for the first positive integer, $n = 1$.
$2^1 = 2 > 1$, so the base case is true.
Step 2: Inductive step
Assume that the statement is true for some positive integer $n = k$.
That is, $2^k > k$.
Now, we need to prove that the statement is also true for $n = k+1$.

$$2^{(k+1)} = 2 * 2^k$$
$$> 2 * k \text{ (by the inductive hypothesis)}$$
$$= 2\,k$$
$$> k + 1 \text{ (since } k > 1)$$

Therefore, the statement is true for all positive integers n.

Example 6.27 Prove that the sum of the first n odd integers is equal to n^2.

Solution: To prove this statement using the Principle of Mathematical Induction, we will follow these steps:

Step 1: Base case
Let us start by verifying the statement for the first positive integer, $n = 1$.
The sum of the first 1 odd integer is 1, which is equal to 1^2.
The base case is true.
Step 2: Inductive step
Assume that the statement is true for some positive integer $n = k$.
That is, the sum of the first k odd integers is equal to k^2.
Now, we need to prove that the statement is also true for $n = k+1$.
The sum of the first $(k+1)$ odd integers is the sum of the first k odd integers plus the $(k+1)$th odd integer, which is $2\,k+1$.
Sum of the first $(k+1)$ odd integers = Sum of the first k odd integers + $(2\,k+1)$

$$= k^2 + (2\,k+1)$$
$$= (k+1)^2$$

Therefore, the statement is true for all positive integers n.

6.6 Factorial Notation and Binomial Coefficients (Combinations)

The binomial theorem is a powerful formula that (among other things) allows us to expand binomial expressions raised to a power in a systematic way. The formula involves a mathematical concept called a "factorial," understanding of which is a crucial prerequisite to comprehension of the theorem.

Factorials are an important concept in mathematics, particularly in combinatorics, probability, and other areas of discrete mathematics. A factorial is the product of all positive integers less than or equal to a given positive integer.

The factorial of a positive integer n is denoted as $n!$ and is calculated as follows:

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$$

For example:

- $1! = 1$
- $2! = 2 \times 1 = 2$
- $3! = 3 \times 2 \times 1 = 6$
- $4! = 4 \times 3 \times 2 \times 1 = 24$
- $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

Combinations, also known as binomial coefficients, are a fundamental concept in mathematics, particularly in the field of combinatorics. Combinations represent the number of ways to choose a certain number of items from a set, without regard to order.

Formally, the combination of k items from a set of n items is denoted as:

$$C(n,k) = n!/(k! * (n-k)!)$$

where

- n is the total number of items in the set.
- k is the number of items to be chosen.
- $n!$ is the factorial of n, which is the product of all positive integers less than or equal to n.

Some key properties and applications of combinations:

1) Counting the number of ways to choose a subset: C(n, k) represents the number of ways to choose k items from a set of n items, without regard to the order of the items.
2) Pascal's triangle: Combinations are closely related to Pascal's triangle, where each number in the triangle is the sum of the two numbers directly above it.
3) Binomial theorem: Combinations are used in the binomial theorem, which expresses the expansion of the binomial $(a + b)^n$ in terms of the coefficients C(n, k).
4) Probability: Combinations are used in probability calculations, such as the probability of choosing a specific subset of items from a set.
5) Permutations and combinations: While permutations consider the order of the chosen items, combinations do not.

Example 6.28 How many ways can you choose three items from a set of five items?

$$C(5,3) = 5!/(3! * (5-3)!) = 10$$

Example 6.29 How many ways can you choose two items from a set of seven items?

$$C(7,2) = 7!/(2! * (7-2)!) = 21$$

Example 6.30 How many ways can you choose 4 items from a set of 10 items?

$$C(10,4) = 10!/(4! * (10-4)!) = 210$$

Example 6.31 Suppose you flip a coin five times. How many different sequences of heads (H) and tails (T) are possible?

To solve this, we need to find the number of combinations of three heads and two tails out of five coin flips.

The formula is: $C(n, k) = n! / (k! * (n-k)!)$

In this case, $n = 5$ (total number of coin flips) and $k = 3$ (number of heads).

$$C(5,3) = 5!/(3! * (5-3)!) = 10$$

Therefore, there are 10 different sequences of heads and tails possible.

Example 6.32 Poker Hands

In a standard deck of 52 playing cards, how many different five-card poker hands are possible?

To find this, we need to calculate the number of combinations of five cards chosen from a set of 52 cards.

$$C(52,5) = 52!/(5! * (52-5)!) = 2{,}598{,}960$$

So, there are 2,598,960 possible five-card poker hands.

Example 6.33 A committee of four people needs to be selected from a group of 10 people. How many different committees are possible?

In this case, we need to find the number of combinations of four people chosen from a group of 10.

$$C(10,4) = 10!/(4! * (10-4)!) = 210$$

Therefore, there are 210 different committees that can be formed.

Example 6.34 Combination Locks

A combination lock has three wheels, each with 10 digits (0–9). How many possible combinations are there?

To find the number of possible combinations, we need to multiply the number of choices for each wheel:

$$C(10,1) * C(10,1) * C(10,1) = 10 * 10 * 10 = 1000$$

So, there are 1000 possible combinations for a three-wheel combination lock.

6.6.1 The Binomial Theorem

The binomial theorem provides a way to expand the expression $(a + b)^n$, where n is a non-negative integer. An expression involving a sum of two variable-containing terms is referred to as a binomial.

The formula for the binomial theorem is:

$$(\mathrm{a}+\mathrm{b})^{\mathrm{n}}=\sum_{\mathrm{k}=0}^{\mathrm{n}} \mathrm{C}(n,k) * a(n-k) * b^{k}$$

where

- n is the exponent of the binomial expression.
- k is the index of the summation, which ranges from 0 to n.
- C(n, k) is the combination of n things taken k at a time, which can be calculated as: C(n, k) = n! / (k! * (n−k)!)
- a and b are the two terms in the binomial expression.

Let us go through some examples to demonstrate the use of the binomial theorem formula.

Example 6.35 Expand $(x+y)^4$

Using the binomial theorem formula:

$$\sum_{\mathrm{k}=0}^{4}(x+y)^4=\mathrm{C}(4,k) * x^{(4-k)*}\, y^{k}$$

Calculating the individual terms:

$k = 0$: $\mathrm{C}(4,0) * x^{4*}y^{0} = 1 * x^{4*}\, 1 = x^4$

$k = 1$: $\mathrm{C}(4,1) * x^{3*}y^{1} = 4 * x^{3*}y = 4x^3y$

$k = 2$: $\mathrm{C}(4,2) * x^{2*}y^{2} = 6 * x^{2*}y^2 = 6x^2y^2$

$k = 3$: $\mathrm{C}(4,3) * x^{1*}y^{3} = 4 * x * y^3 = 4xy^3$

$k = 4$: $\mathrm{C}(4,4) * x^{0*}y^{4} = 1 * 1 * y^4 = y^4$

Putting it all together:

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

Example 6.36 Expand $(2a-3b)^5$

$$\sum_{\mathrm{k}=0}^{5}$$

Using the binomial theorem formula:

$$(2a-3b)^5 = \mathrm{C}(5,k) * (2a)^{(5-k)*}\,(-3b)^{k}$$

Calculating the individual terms:

$k = 0$: $\mathrm{C}(5,0) * (2a)^{5*}\,(-3b)^{0} = 1 * 32a^{5*}\, 1 = 32a^5$

$k = 1$: $\mathrm{C}(5,1) * (2a)^{4*}\,(-3b)^{1} = 5 * 16a^{4*}\,(-3b) = -240a^4b$

$k = 2$: $\mathrm{C}(5,2) * (2a)^{3*}\,(-3b)^{2} = 10 * 8a^{3*}\, 9b^2 = 720a^3b^2$

$k = 3$: $\mathrm{C}(5,3) * (2a)^{2*}\,(-3b)^{3} = 10 * 4a^{2*}\,(-27b^3) = -1080a^2b^3$

$k = 4$: $\mathrm{C}(5,4) * (2a)^{1*}\,(-3b)^{4} = 5 * 2a * 81b^4 = 810ab^4$

$k = 5$: $\mathrm{C}(5,5) * (2a)^{0*}\,(-3b)^{5} = 1 * 1 * (-243b^5) = -243b^5$

Putting it all together:

$$(2a - 3b)^5 = 32a^5 - 240a^4b + 720a^3b^2 - 1080a^2b^3 + 810ab^4 - 243b^5$$

These examples demonstrate how to use the binomial theorem formula to expand binomial expressions. The key steps are:

1) Identify the values of n and a, b in the expression.
2) Calculate the combination $C(n, k)$ for each term.
3) Multiply the appropriate powers of a and b for each term.
4) Add up all the terms to get the final expanded expression.

This approach can be applied to any binomial expression, without the need for programming.

Example 6.37 Expand $(3x - 2y)^6$

Using the binomial theorem formula:

$$\sum_{k=0}^{6} (3x - 2y)^6 = C(6, k) * (3x)^{(6-k)*} (-2y)^k$$

Calculating the individual terms:

$k = 0$: $C(6,0) * (3x)^{6*} (-2y)^0 = 1 * 729x^{6*} 1 = 729x^6$

$k = 1$: $C(6,1) * (3x)^{5*} (-2y)^1 = 6 * 243x^{5*} (-2y) = -1458x^5y$

$k = 2$: $C(6,2) * (3x)^{4*} (-2y)^2 = 15 * 81x^{4*} 4y^2 = 2430x^4y^2$

$k = 3$: $C(6,3) * (3x)^{3*} (-2y)^3 = 20 * 27x^{3*} (-8y^3) = -4320x^3y^3$

$k = 4$: $C(6,4) * (3x)^{2*} (-2y)^4 = 15 * 9x^{2*} 16y^4 = 2160x^2y^4$

$k = 5$: $C(6,5) * (3x)^{1*} (-2y)^5 = 6 * 3x * (-32y^5) = -576xy^5$

$k = 6$: $C(6,6) * (3x)^{0*} (-2y)^6 = 1 * 1 * 64y^6 = 64y^6$

Putting it all together:

$$(3x - 2y)^6 = 729x^6 - 1458x^5y + 2430x^4y^2 - 4320x^3y^3 + 2160x^2y^4 - 576xy^5 + 64y^6$$

Example 6.38 Expand $(a + b)^7$

$$\sum_{k=0}^{7}$$

Using the binomial theorem formula:

$$(a + b)^7 = C(7, k) * a^{(7-k)*} b^k$$

Calculating the individual terms:

$k = 0$: $C(7,0) * a^{7*} b^0 = 1 * a^{7*} 1 = a^7$

$k = 1$: $C(7,1) * a^{6*} b^1 = 7 * a^{6*} b = 7a^6b$

$k = 2$: $C(7,2) * a^{5*} b^2 = 21 * a^{5*} b^2 = 21a^5b^2$

$k = 3$: $C(7,3) * a^{4*} b^3 = 35 * a^{4*} b^3 = 35a^4b^3$

$k = 4$: $C(7,4) * a^{3*} b^4 = 35 * a^{3*} b^4 = 35a^3b^4$

$k = 5$: C(7,5) $* a^{2*}b^5 = 21 * a^{2*}b^5 = 21a^2b^5$

$k = 6$: C(7,6) $* a^{1*}b^6 = 7 * a * b^6 = 7ab^6$

$k = 7$: C(7,7) $* a^{0*}b^7 = 1 * 1 * b^7 = b^7$

Putting it all together:

$$(a+b)^7 = a^7 + 7a^6b + 21a^5b^2 + 35a^4b^3 + 35a^3b^4 + 21a^2b^5 + 7ab^6 + b^7$$

6.6.2 Graphical Interpretation and Pascal's Triangle

We can visualize the binomial theorem using a graph of the binomial coefficients. The graph below shows the binomial coefficients for $n = 0$ to 6, where each row represents the coefficients for a given value of n.

```
          1
        1   1
      1   2   1
    1   3   3   1
  1   4   6   4   1
1   5   10  10  5   1
1   6   15  20  15  6   1
```

The pattern of the binomial coefficients forms a triangle, known as Pascal's triangle. This triangle has many interesting properties and is closely related to the binomial theorem.

6.6.3 Applications of the Binomial Theorem

The binomial theorem has numerous applications in mathematics, including:

1) Binomial probabilities: The binomial theorem can be used to calculate the probabilities of events in a binomial probability distribution.
2) Taylor series expansion: The binomial theorem is the basis for the expansion of functions using Taylor series, which is a powerful tool in calculus and mathematical analysis.
3) Combinatorial identities: The binomial theorem leads to many important combinatorial identities, such as the formula for the sum of the first n positive integers.
4) Algebra and number theory: The binomial theorem has applications in various areas of algebra and number theory, such as the expansion of binomial expressions and the study of binomial coefficients.

6.6.4 Applications of the Binomial Theorem in Cryptography

The binomial theorem has numerous applications in the field of cryptography, which is the study of secure communication techniques. In this section, we will explore two important ways in which the binomial theorem is utilized in cryptographic systems.

6.6.4.1 Modular Arithmetic and the Binomial Theorem

One of the key applications of the binomial theorem in cryptography is in the context of modular arithmetic. Modular arithmetic is the study of the behavior of numbers under the operation of division with a fixed divisor, known as the modulus.

Consider the expression $(a + b)^n$ (mod m), where a, b, n, and m are all integers. Using the binomial theorem, we can expand this expression as:

$$(a + b)^n \text{ (mod m)} = \Sigma \, (n\text{Cr}) \, a^{(n-\text{r})} \, b^r \text{ (mod m)}$$

This formula allows us to efficiently compute large powers of binomial expressions modulo a given modulus, which is an important operation in many cryptographic algorithms.

For example, the RSA cryptosystem, one of the most widely used public-key cryptographic algorithms, relies heavily on modular arithmetic and the binomial theorem. The RSA algorithm involves raising a message to a power (the public or private key) modulo a large composite number. The binomial theorem provides an efficient way to perform these computations.

6.6.4.2 The Binomial Theorem and Pseudo-random Number Generation

Another application of the binomial theorem in cryptography is in the generation of pseudo-random numbers. Pseudo-random number generators (PRNGs) are algorithms that produce sequences of numbers that appear to be random but are actually deterministic.

One class of PRNG algorithms, known as linear congruential generators (LCGs), uses the binomial theorem to generate their output. The basic idea behind an LCG is to generate a sequence of numbers using the recurrence relation:

$$\text{X}_{n+1} = (\text{a} * \text{X}_n + \text{b}) \text{ (mod m)}$$

where a, b, and m are carefully chosen constants.

The binomial theorem can be used to analyze the properties of the sequence generated by an LCG, such as its period and statistical characteristics. This understanding helps cryptographers design more secure and efficient pseudo-random number generators, which are essential components of many cryptographic systems.

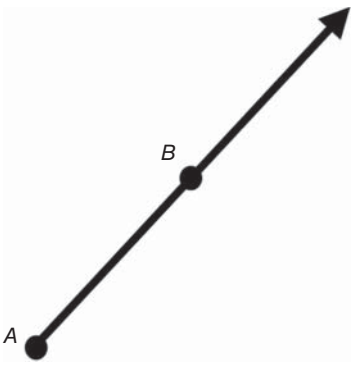
7

Measuring an Angle

7.1 Measuring an Angle

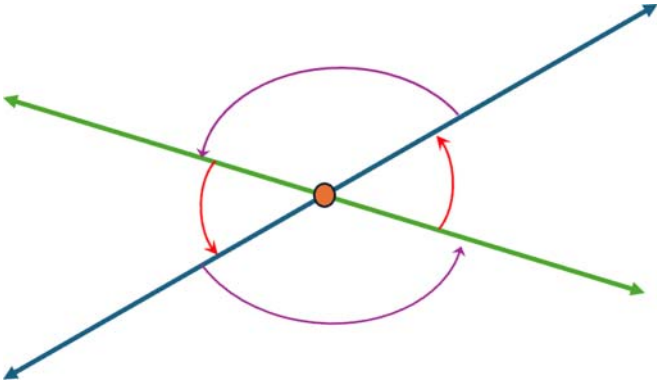
Angles are a fundamental concept in geometry and trigonometry. Before we can understand that notion, we must recall the definition of a ray in the plane; this is a building block of the idea of an angle, so it is appropriate to begin there.

A ray in the plane is equivalent to a half-line in the plane. There is a starting point, or initial point, to the ray, and the points on the ray extend in a straight-line path from that initial point endlessly in one direction. We describe a ray by specifying its initial point A, a second point on the ray B, and using the notation $\vec{AB}$. The point B is NOT the terminal point of the ray, because a ray has no terminal point.



An angle is the amount of rotation between two intersecting lines or rays in a plane. Angles are typically measured in degrees (°) or radians (rad). When considered as the amount of rotation between two rays, those rays are positioned in such a way that their initial points coincide. When looked upon as the amount of rotation between two intersecting lines, the point of intersection is viewed as the initial point, and the half-lines created by breaking the lines at this point form four rays between which angles can be identified.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

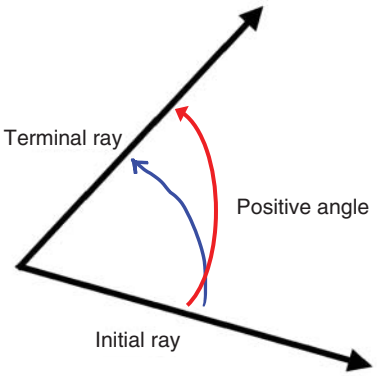


Note that if intersecting lines are viewed as forming an "X," the angles not sharing terminal sides are called "vertical angles" to one another, and a fundamental fact is that vertical angles are always equal to one another.

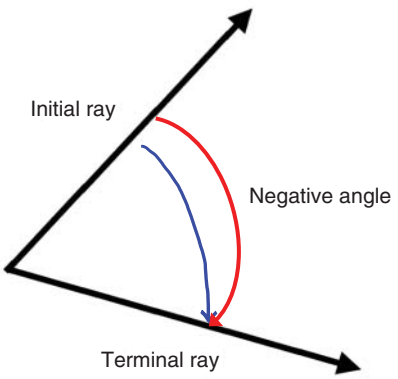
For the purpose of reference, one of the rays is selected to be the initial ray, the other is (by default) the terminal ray of the angle. If the initial ray were to coincide with the positive x-axis and the initial point of that ray to be the origin, we say the initial ray is in "standard position." This is a useful choice of initial ray, as much of the geometry of angles then has a suitable source of reference.

By general agreement, counterclockwise (or anticlockwise) rotation from the initial ray is considered positive, and clockwise rotation is considered negative.

Positive Angle



Negative Angle



If the two rays are positioned so the initial point lies at the origin and the initial side lies along the positive x-axis, we say the angle is drawn in standard position.

We saw the word "standard" in the context of linear equations, and at that time, we stated that the notion of a standard form for a linear equation served the purpose of allowing us to determine if different forms of the equation of a line represented the same line. In the context of angles, the concept is the same: angles can be drawn anywhere in the plane, but to determine if they are equivalent, we shift their location to standard position, from which we can establish if the angles are equivalent.

7.1.1 Degrees

As has been said, the measure of an angle is the amount of rotation from the initial ray to the terminal ray. The amount of rotation must be measured in some way, and it is to this that we now turn our attention. We must establish scales by which this rotation can be determined, and there are two methods that are typically used.

The most common unit of angle measurement is the degree (°). A full rotation, or circle, is equal to 360 degrees. Each degree is further divided into 60 minutes (‘), and each minute is divided into 60 seconds (″).

For example, an angle of $37°24'15''$ consists of:

- 37 degrees
- 24 minutes
- 15 seconds

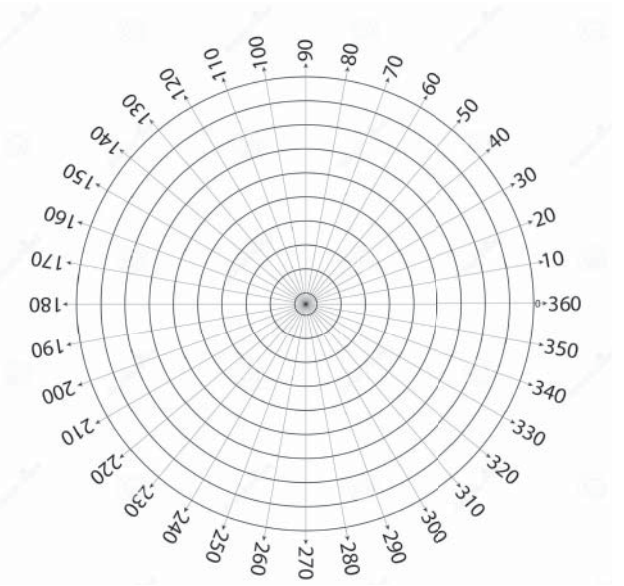
It may seem odd that a complete rotation would be subdivided into 360 degrees. The concept has its origin in antiquity, over 4000 years ago to the culture of Babylon. The Babylonians are responsible for dividing the circle up into 360 equal parts, which we call "degrees," likely due to the length of the year being approximately 365 days, and the observation that in 1 day, the Sun moves about 1° with respect to the stars.

In a year, therefore, the sun would have moved a total of approximately 360°—and in doing so, had returned to its original position. In a sense, you might say the Earth had come *full circle* about the Sun. This observation, we remark, is not quite correct (1 year is not 360 days long), and this led to the errors within calendars, which had to be corrected centuries later.

The notion that the passage of time is related to this manner of subdivision of a circle into 360° is supported, in part, by the further subdivision of degrees into minutes and seconds (also units of time). Typically, the terminology used in any endeavor has a logical basis, and understanding this can help you develop command over the concepts being described.

Let us consider the subdivision of angles into degrees, beginning with the understanding that an entire revolution around a circle is 360°. As this is positive, we keep in mind that we are discussing a counterclockwise rotation. From this presupposition, one-half of a revolution about a circle must be 180°, and one-quarter of a revolution about a circle must be 90°.

Degree Measure of Angles about a Central Point

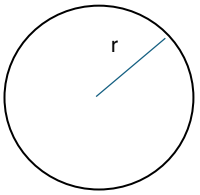


A 90° angle is referred to as a right angle. Angles that are greater than 0° yet less than 90° are called acute, while angles between 90° and 180° are called obtuse.

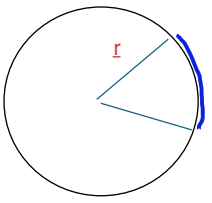
7.1.2 Radians

Radians (rad) are an alternative unit of angle measurement, where the rotation about a full circle is equal to 2π radians. One radian is the angle that subtends an arc length equal to the radius of the circle.

To illustrate this, imagine a circle of arbitrary radius r:



If a flexible cord of radius r were wrapped about the circle and two radial lines drawn to the ends of that flexible cord, the radial lines would "subtend" or "create" an angle between them.



This angle is, by definition, one radian in length.

Observe that, had the radius of the circle been precisely 1 unit, the circumference of the circle (given by the formula $2\pi r$) tells us one rotation about the center of the circle is 2π radians.

Remember, despite the different systems of measurement, an angle is an angle. Therefore, the agreement that one full rotation about the center is 360°, and also 2π radians tell us there is an equivalence relationship present:

$$360^\circ = 2\pi \text{ radians}$$

We can, thus, find the equivalence in degrees to 1 radian by dividing this equation by 2π.

The relationship between degrees and radians is:

1) 1 radian $\approx$ 57.296 degrees
2) 1 degree $\approx$ 0.01745 radians

Radians are often preferred in mathematical and scientific applications because they have a more direct relationship with other trigonometric functions and because they are "unitless" measures.

Radians are considered unitless because they are essentially a ratio of two lengths (the arc length of a circle to its radius), which mathematically cancels out the units, leaving only a pure number representing the angle; in other words, a radian is a "dimensionless" unit, meaning it doesn't have a unit attached to it like meters or seconds. Degrees, on the other hand, are considered to have units because they represent a standardized way of measuring angles, just like meters measure length or seconds measure time. Even though "degree" is not part of the International System of Units (SI), it is still a defined unit used to quantify an angular measurement. When performing computations in applied problems, the issue of units of measure becomes important when attempting to produce answers having specific, necessary final units.

7.1.3 Conversion Between Degrees and Radians

It frequently occurs that we must convert between the two units of angle measure. Doing so requires the use of a "conversion factor." A conversion factor (sometimes called a scaling factor) is a fraction equivalent to the number 1, where the numerator and denominator describe the same measure but use different scales. Anytime we have an equivalence of measures, we can create two conversion factors.

Example 7.1 It is well-known that 1 yard = 36 inches. What are the two conversion factors associated with this equivalence?

Solution:

A conversion factor is a fraction equivalent to the number 1, where the numerator and denominator describe the same measure but use different scales. In this case, our equivalence produces the conversion factors:

$$\frac{1\ yard}{36\ inches} \text{ and } \frac{36\ inches}{1\ yard}$$

Example 7.2 1 week is equivalent to 7 days. What are the two conversion factors associated with this equivalence?

Solution:
Again, a conversion factor is a fraction equivalent to the number 1, where the numerator and denominator describe the same measure but use different scales. In this case, our equivalence produces the conversion factors:

$$\frac{1\ week}{7\ days} \quad and \quad \frac{7\ days}{1\ week}$$

Example 7.3 360° is equivalent to 2π radians. What are the two conversion factors associated with this equivalence?

Solution:
Since we have stated 360 degrees is equivalent to 2π radians, the two potential conversion factors between these units of angle measure are:

$$\frac{360°}{2\pi\ radians} \quad and \quad \frac{2\pi\ radians}{360°}$$

The use of conversion factors relative to angle measurement is best illustrated through example, in which we will explain the logic for the selection between the two alternatives. In these conversion factors, the word "radians" may be suppressed if we maintain awareness of that measurement scale.

Example 7.4 Convert 135° to radians.

Solution:
The given measure is first stated as a fraction; if the measure is a whole number, it is expressed as a fraction by using a denominator of 1.

$$\frac{135°}{1}$$

The existing measurement units, degrees, appear in the numerator of this fraction. When using a conversion factor, the selection of which factor to use should have the given units in the denominator and the desired units in the numerator. This factor should be multiplied by the fraction created using the given measure.

$$\frac{135°}{1} \times \frac{2\pi\ radians}{360°}$$

We simplify recognizing there is a common factor of 45 between 135 and 360 and that the units of degrees cancel through this multiplication.

$$\frac{3}{1} \times \frac{2\pi}{8}$$

$$\frac{6\pi}{8}$$

$$\frac{3\pi}{4}$$

Thus, 135° is equivalent to $\frac{3\pi}{4}$ radians.

Example 7.5 Convert 20 radians to degrees, rounding your answer to four decimal places.

Solution:
We create a fraction using the given measure:

$$\frac{20\ radians}{1}$$

The given units are radians, and, therefore, we should use the conversion factor having radians in the denominator.

$$\frac{20\ radians}{1} \times \frac{360^\circ}{2\pi\ radians}$$

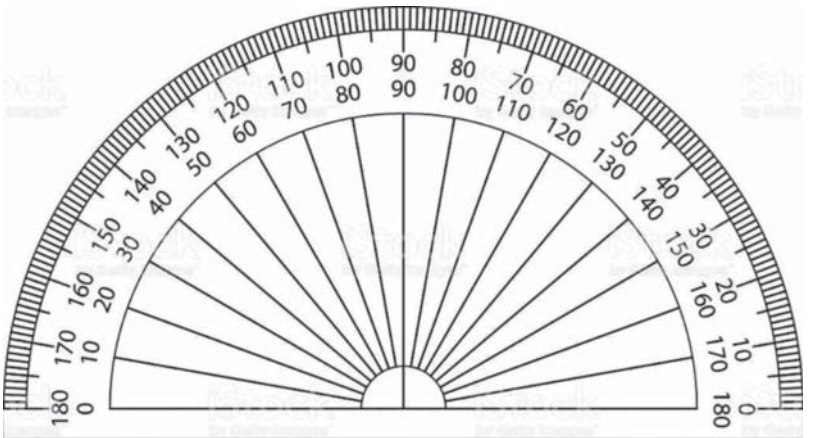
Canceling the units of radians and using the approximation to four decimal places, we find 20 radians is equivalent to:

$$1145.9156^\circ$$

Note that, in the preceding example, we find the immense difference between radians and degrees, and the need for recognizing which of the two measures we are using. When we fail to distinguish between 20 radians and 20°, we are not "close to the truth." We are, as seen above, more than 1000° away from the truth, which is no small distinction.

7.1.4 Measuring Angles

Angles can be measured using a **protractor**, a tool with a semicircular scale marked in degrees.

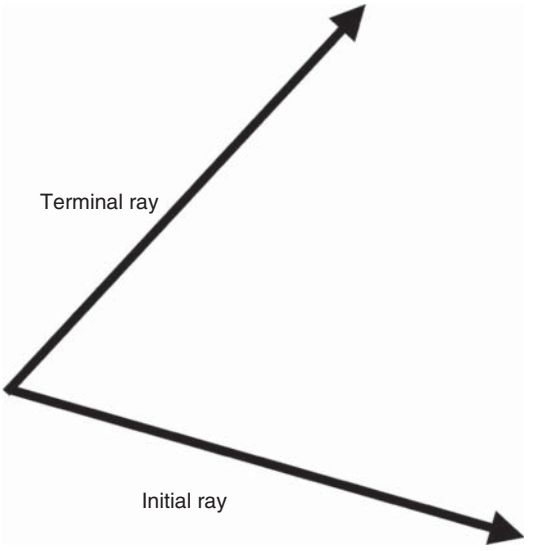


To measure an angle with a protractor:

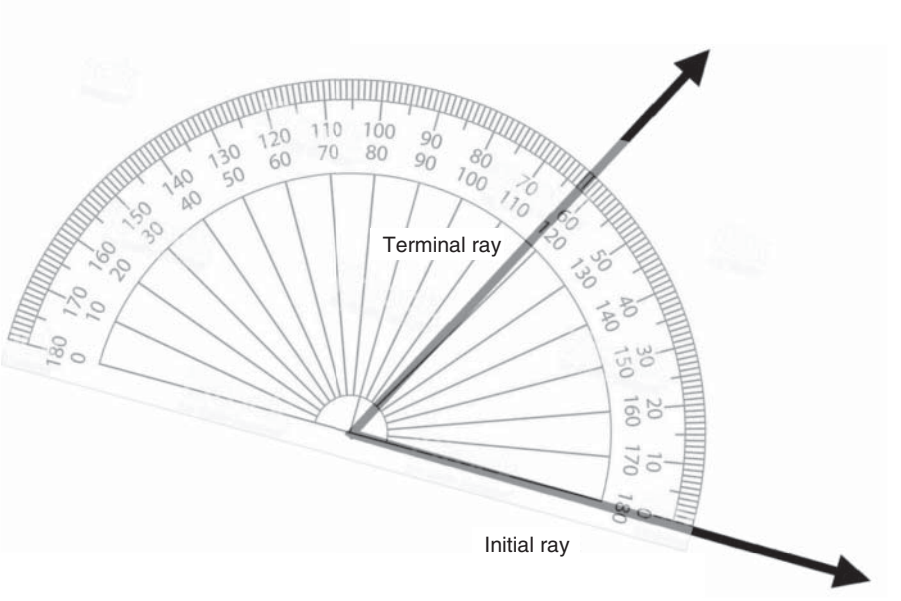
1) Place the protractor so its central point is on the vertex of the angle and the straight line to the right of the central point is aligned with the initial side of the angle.
2) Read the angle measurement from the scale on the protractor; carefully note the counterclockwise (anticlockwise) measure is on the outside of the protractor.

Alternatively, angles can be measured using a compass and straightedge constructions. This involves constructing geometric figures with which we can determine the angle measurement.

Example 7.6 Consider the following angle and determine the positive angle determined by the rays.

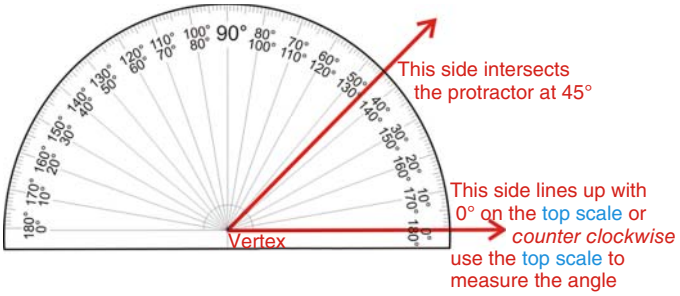


Solution:
We superimpose a protractor upon the image, aligning the central point with the vertex and the horizontal line to the right of that central point with the initial ray.



The angle shown measures 62°.

Example 7.7 Measure the angle shown in the figure below.

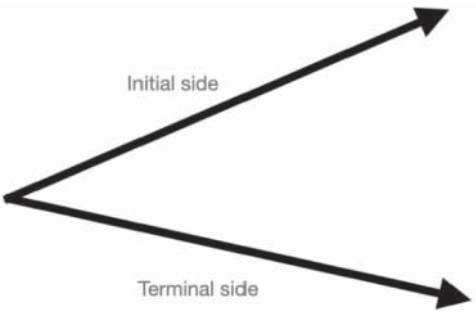


Source: CK-12 Foundation / https://dr282zn36sxxg.cloudfront.net/datastreams/f-d%3A7fee8ad96b220d3c8fb20df1c442939313b321d8b6af3d3494c91587%2BIMAGE_TINY%2BIMAGE_TINY.1.

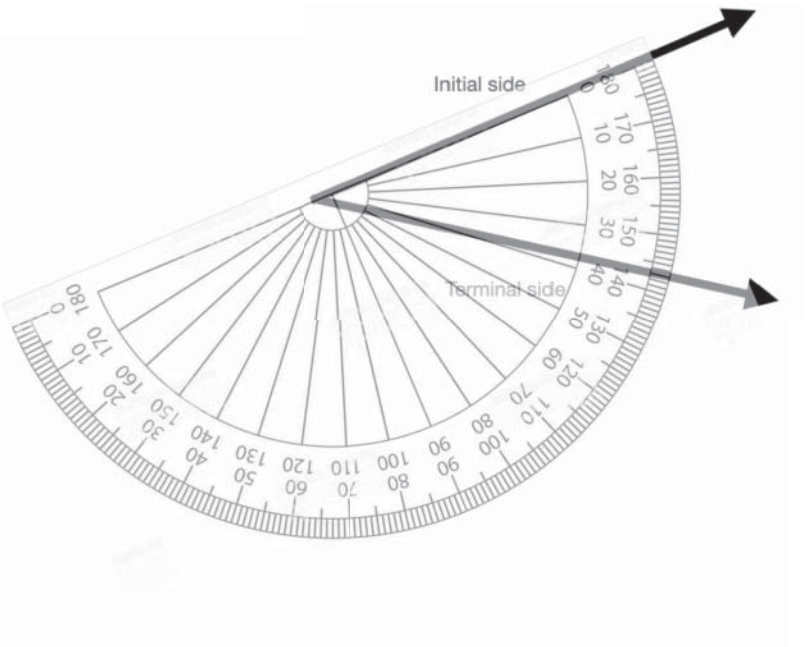
Protractors can be used to measure negative angles, as well. This involves recognizing an angle is negative and using one of two methods for positioning the protractor:

1) We can place the central point of the protractor at the vertex and align the horizontal line to the right of the central point with the TERMINAL point of the angle, reading the result as before. Using our awareness, the angle is negative; the measurement can be changed to a negative value.
2) We can place the central point of the protractor at the vertex and align the reference line to the left of that central point with the initial point of the angle. The angle, again recognized to be negative, is read using the inner value on the ring of angle measures on the protractor.

Example 7.8 Find the measure of the shown angle using a protractor.



Solution:



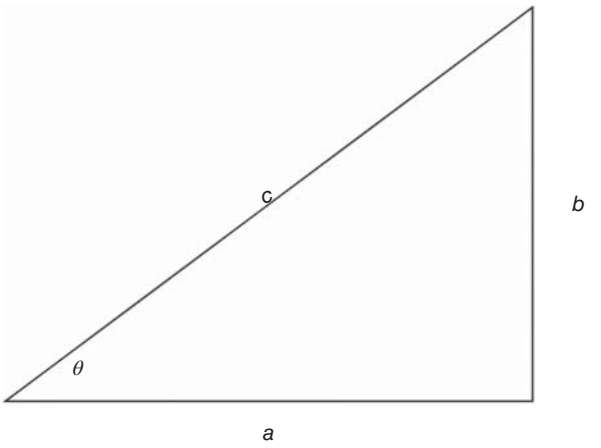
The shown angle is $-36°$.

7.2 Trigonometric Functions

7.2.1 Right Triangles

A right triangle is a triangle with one 90° angle and two acute angles. The side opposite the 90° angle is called the hypotenuse, while the other two sides are called legs of the triangle. One of the non-right angles is labeled or indicated as the angle of interest in a particular situation, and relative to this identification, the leg meeting the hypotenuse at the specified angle is referred to as the adjacent side, while the remaining leg is referred to as the opposite side.

The hypotenuse is always the longest side of the right triangle. The three sides of the triangle are related to one another using the Pythagorean theorem.



In the figure above, the hypotenuse is labeled as "c," the adjacent side is "a," and the opposite side is "b." The angle θ is the angle opposite the side "b."

Using these conventions, the Pythagorean theorem tells us:

$$a^2 + b^2 = c^2$$

Trigonometric functions capture fundamental relationships between the sides and angles of a right triangle. The three primary trigonometric functions are sine (sin), cosine (cos), and tangent (tan). These functions are given in reference to one of the acute angles of the right triangle.

Using the identifications mentioned above, we define the three primary trigonometric functions of an acute angle of a right triangle.

7.2.2 Primary Trigonometric Functions

The three primary trigonometric functions are defined as follows:

1) $\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}$; that is, the sine of an acute angle of a right triangle is the ratio of the opposite side to the hypotenuse.
2) $\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}$; that is, the cosine of an acute angle of a right triangle is the ratio of the adjacent side to the hypotenuse.
3) $\tan\theta = \frac{\text{opposite}}{\text{adjacent}}$; that is, the tangent of an acute angle of a right triangle is the ratio of the opposite side to the adjacent side.

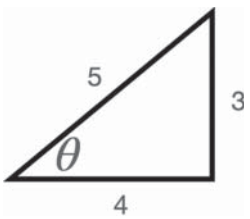
The definition of these functions can be remembered using the mnemonic device "SOH-CAH-TOA":

- Sine = Opposite/Hypotenuse.
- Cosine = Adjacent/Hypotenuse.
- Tangent = Opposite/Adjacent.

Example 7.9 Find the values of $\sin(\theta)$, $\cos(\theta)$, and $\tan(\theta)$ for a right triangle with sides $a = 3$, $b = 4$, and $c = 5$. The angle θ lies at the intersection of the sides having lengths 4 and 5.

Solution:
We have a right triangle with sides $a = 3$, b = 4, and $c = 5$. The hypotenuse is the longest of these sides, in this case $c = 5$.

Recognizing the hypotenuse is 5 and being aware the angle θ lies at the intersection of the hypotenuse and the leg having length 4, we visualize the situation as follows:



Observe, therefore:

Opposite side = 3
Adjacent side = 4
Hypotenuse = 5

Plugging these values into the trigonometric ratio formulas:

$\sin(\theta) = 3/5$
$\cos(\theta) = 4/5$
$\tan(\theta) = 3/4$.

7.2.3 Secondary Trigonometric Functions

We have discussed sine, cosine, and tangent to this point; there are three additional trigonometric functions, each of which is the reciprocal of one of the aforementioned functions.

The Cosecant Function:

$$\csc\theta = \frac{1}{\sin\theta}$$

The Secant Function:

$$\sec\theta = \frac{1}{\cos\theta}$$

The Cotangent Function:

$$\cot\theta = \frac{1}{\tan\theta}$$

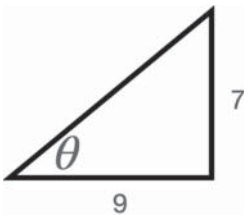
You may notice the recurrence of the prefix "co" in half of the listed trigonometric functions. We refer to the functions whose names involve "co" as "co-functions." That is, for sine, there is the co-function cosine. For secant, there is the co-function cosecant. For tangent, there is the co-function cotangent.

More will be said of these shortly.

7.2.4 Simplifications That Are Not Exact

The Pythagorean theorem is one of the most important identities in all of geometry, and we have seen its use in computations pertaining to the trigonometric functions of an acute angle within a right triangle. We would be remiss, however, in presenting only examples where the sides of the triangles are whole numbers.

Consider the right triangle having legs measuring 7 and 9 inches, as shown:



The hypotenuse can be found by using the Pythagorean theorem, as we have already demonstrated. We can use the identification $a = 9$ and $b = 7$, with the as-yet unknown hypotenuse identified as c.

$9^2 + 7^2 = c^2$
$81 + 49 = c^2$

$$130 = c^2$$

$$\sqrt{130} = c$$

Observe we have not used the ± symbol here; the value of c is the length of the hypotenuse and therefore is positive.

Unless asked, it would be inappropriate for us to shift to a decimal approximation at this point. Using the methods of college algebra, we would need to algebraically simplify this square root, if possible. If simplification is not possible, then (unappealing as the value may look) we leave it in the radical form.

The factorization of 130 into primes is $2 \times 5 \times 13$, and therefore, there are no perfect-square factors of the radicand 130, so $\sqrt{130}$ cannot be simplified and is the length of the hypotenuse.

7.2.5 Recognizing When a Square Root Can Be Simplified and Performing the Simplification

As an illustration, we regard the square root of 75, $\sqrt{75}$. Square roots "simplify" if we can find a perfect square that divides into the radicand evenly. Recall the perfect squares are:

4, 9, 16, 25, 36, 49, 64, 81, 100, 121, and so on.

From this list, we should locate the largest perfect square (if any) that is a factor of the radicand. Here, observe this is 25. We use that value to produce a factorization of 75 as 25×3, so our expression becomes

$$\sqrt{75} = \sqrt{(25)(3)} = \sqrt{25}\sqrt{3}$$

Observe the first square root in the product simplifies, and we have

$$\sqrt{75} = 5\sqrt{3}$$

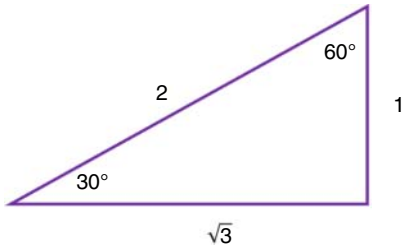
No matter how many decimal places of accuracy are used, no approximation suffices as the simplified form of $\sqrt{75}$!

7.2.6 Fundamental Right Triangles to Be Memorized

There are two fundamental right triangles we must commit to memory, leading to the exact values of the trigonometric functions of three specific angles, 30°, 45°, and 60°. The reason these triangles are significant is that they appear in many applications, and their exact trigonometric function values are used significantly.

7.2.6.1 The $1-2-\sqrt{3}$ Triangle

All right triangles having acute angles 30° and 60° are geometrically similar (proportional) to the following triangle:



Consideration of this triangle (and appropriate labeling of adjacent, opposite, and hypotenuse sides) gives us the following "known measures":

$$\sin 30^\circ = \frac{1}{2} = \cos 60^\circ$$

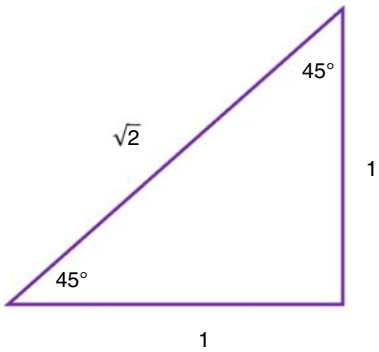
$$\cos 30^\circ = \frac{\sqrt{3}}{2} = \sin 60^\circ$$

$$\tan 30^\circ = \frac{\sqrt{3}}{3}$$

$$\tan 60^\circ = \sqrt{3}$$

7.2.6.2 The $1-1-\sqrt{2}$ Triangle

All right triangles having acute angles 45° and 45° are geometrically similar (proportional) to the following triangle:



Consideration of this triangle (and appropriate labeling of adjacent, opposite, and hypotenuse sides) gives us the following "known measures":

$$\sin 45^\circ = \frac{\sqrt{2}}{2} = \cos 45^\circ$$

$$\tan 45^\circ = 1$$

Example 7.10 Find the lengths of the adjacent and opposite sides in a right triangle with a hypotenuse of 10 units and an acute angle $\theta = 30$ degrees.

Solution:

We know that for the acute angle 30 degrees, we have the following ratios for the trigonometric functions:

$$\sin(30°) = 1/2$$
$$\cos(30°) = \sqrt{3}/2$$
$$\tan(30°) = \frac{\sqrt{3}}{3}$$

We use the definitions of the trigonometric functions, plugging in the hypotenuse of 10 units:

$$\frac{1}{2} = \frac{\text{opposite}}{10}$$

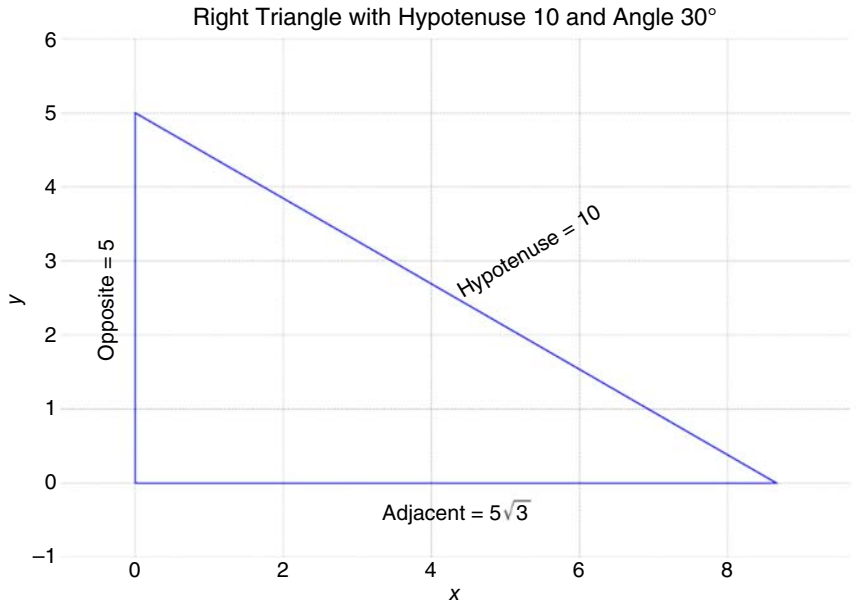
Multiplication of each side of this equation produces

$$\frac{10}{2} = \text{opposite}$$
$$5 = \text{opposite}$$

The measure of the adjacent side can be found using the cosine function, or we could use the Pythagorean theorem. We will present the latter choice:

$$5^2 + a^2 = 10^2$$
$$a^2 = 75$$
$$a = \sqrt{75}$$
$$a = 5\sqrt{3}$$

Therefore, the length of the opposite side is 5 units, and the length of the adjacent side is 8.66 units.



7.2.7 Trigonometric Function Values of Any Real Number and Graphing a Trigonometric Function

To this point, we have looked at trigonometric functions defined within the context of a right triangle. However, if we were to ask for sin 20, what would that mean? Remember, without a degree symbol visible, this is 20 radians. We showed earlier this was equivalent to 1145.9156 degrees—clearly this is not an angle in a right triangle, so what are we discussing?

Keep in mind 2π radians correspond to a single revolution about the plane. Therefore, 20 radians is approximately $\frac{20}{2\pi} \approx 3.18$ revolutions about the plane. This culminates at a terminal side of the angle, and the sine of this angle is equivalent to the sine of the acute angle having that same terminal side; we call such angles coterminal.

Note the coterminal angle is approximately 69.9156° and that (using a calculator) we see

$$\sin 1145.9156^\circ \approx 0.9129 \approx \sin 69.9156^\circ$$

Observe, furthermore, that

$$\sin 1145.9156^\circ = \sin 785.9156^\circ = \sin 425.9156^\circ = \sin 69.9156^\circ$$

In other words, the values of sine are repeating every 360°; they are "periodically" repeating.

This suggests to us that, were we to graph the sine function over a $360^\circ = 2\pi$ interval, the graph would thereafter repeat itself infinitely in both directions along the x-axis.

Functions in two variables can be graphed in the Cartesian plane, using one among several methods. Perhaps the most direct method of finding the graph of a function is by plotting points from a table of values, and we will use this method to examine the graphs of trigonometric functions.

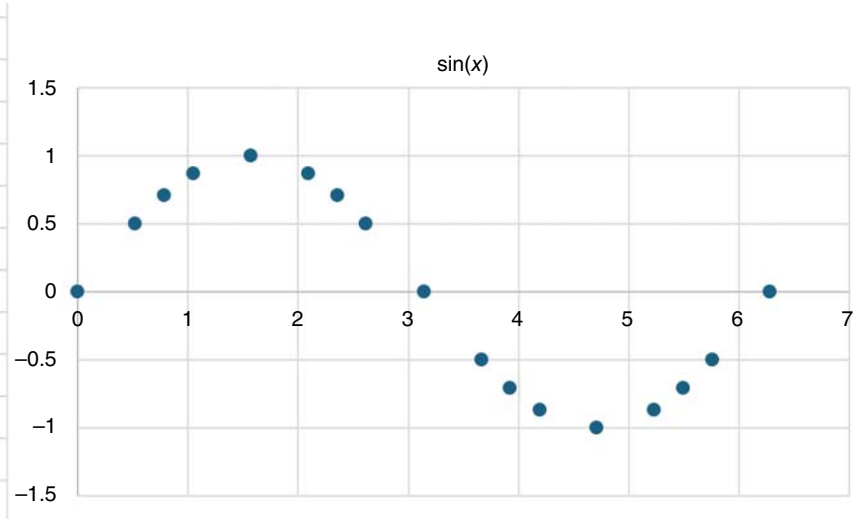
It is important to realize that the angles used in graphing the trigonometric functions are always measured in radians. This is by general agreement, and we must adhere to this requirement. For the purposes of constructing a table of values, we will calculate decimal approximations to the trigonometric function values, as those will be more informative than the exact representations of the values.

x	$\sin x$
0	
$\pi/6$	
$\pi/4$	
$\pi/3$	
$\pi/2$	
$2\pi/3$	
$3\pi/4$	
$5\pi/6$	
π	
$7\pi/6$	
$5\pi/4$	
$4\pi/3$	
$3\pi/2$	
$5\pi/3$	
$7\pi/4$	
$11\pi/6$	
2π	

If we were to now calculate the y-coordinates of these points, we would find (decimal approximating, for visual convenience):

x	$\sin x$
0	0
$\pi/6$	0.5
$\pi/4$	0.71
$\pi/3$	0.87
$\pi/2$	1
$2\pi/3$	0.87
$3\pi/4$	0.71
$5\pi/6$	0.5
π	0
$7\pi/6$	−0.5
$5\pi/4$	−0.71
$4\pi/3$	−0.87
$3\pi/2$	−1
$5\pi/3$	−0.87
$7\pi/4$	−0.71
$11\pi/6$	−0.5
2π	0

Plotting these points, we find:



Notice how the points appear to follow an up-and-down curve; we call this path a sinusoidal curve. This is the "pattern" the sine function will follow forever to the left and to the right. Observe

the highest the curve will reach is "1" on the y-axis, the lowest the curve will reach is "-1" on the y-axis.

Draw a smooth curve passing through these points and repeat this pattern endlessly in both directions, and you will obtain the graph of a sine function.

Example 7.11 Sketch the graph of $y = \sin(x + \pi/4)$.

To sketch the graph, we can use our knowledge of the properties of the sine function and the effect of shifting the input by $\pi/4$.

The sine function has the following properties:

Period: 2π
Amplitude: 1
Range: $[-1, 1]$

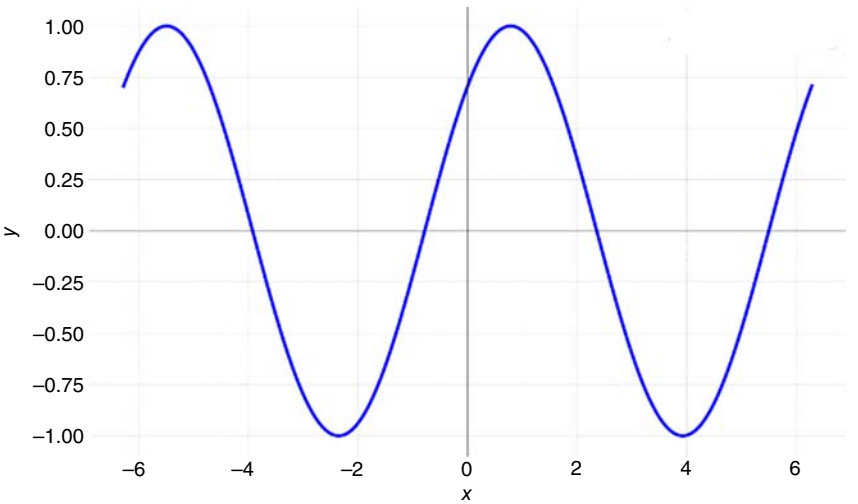
Now, let us consider the effect of the shift by $\pi/4$:

The input to the sine function is now $x + \pi/4$ instead of just x.

This means the graph of $y = \sin(x + \pi/4)$ will be shifted to the left by $\pi/4$ compared to the graph of $y = \sin(x)$.

To sketch the graph, we can start by sketching the graph of $y = \sin(x)$, which is a sine wave with period 2π and amplitude 1.

Then, we can shift the graph to the left by $\pi/4$. This will result in the following graph:



This graph demonstrates how the sum and difference identities can be used to transform and manipulate trigonometric functions, which is a crucial skill for solving a wide range of trigonometric problems.The key points on the graph are:

At $x = -\pi/4$, the function has a value of 0.
At $x = 0$, the function has a value of $\sqrt{2}/2$.
At $x = \pi/4$, the function has a value of 1.
At $x = \pi$, the function has a value of 0.

So this is the sketch of the graph of $y = \sin(x + \pi/4)$.

Example 7.12 Sketch the graph of $y = \cos(x - \pi/3)$.

To sketch the graph, we need to understand the properties of the cosine function and how the shift by -π/3 affects the graph.

The properties of the cosine function:

The cosine function is periodic with a period of 2π.
The cosine function oscillates between −1 and 1, with a maximum at 0 and minimums at π and 2π.
The cosine function is even, meaning $\cos(-x) = \cos(x)$.

Now, the shift by $-\pi/3$:

Shifting the function to the right by $\pi/3$ means that the function is now centered around $x = \pi/3$ instead of $x = 0$.

This means that the maximum of the function will now occur at $x = \pi/3$, and the minimums will occur at $x = 4\pi/3$ and $7\pi/3$.

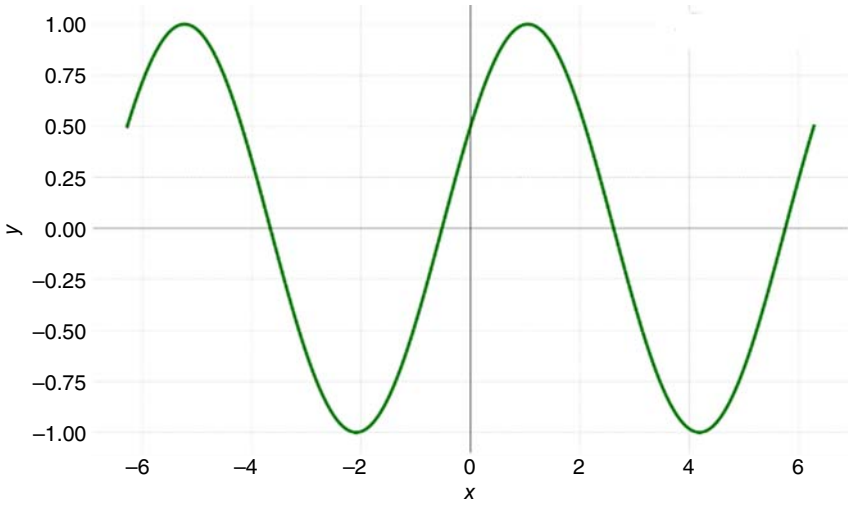
To sketch the graph:

Draw the basic cosine function, $y = \cos(x)$, with its period and range.

Shift the graph to the right by $\pi/3$.

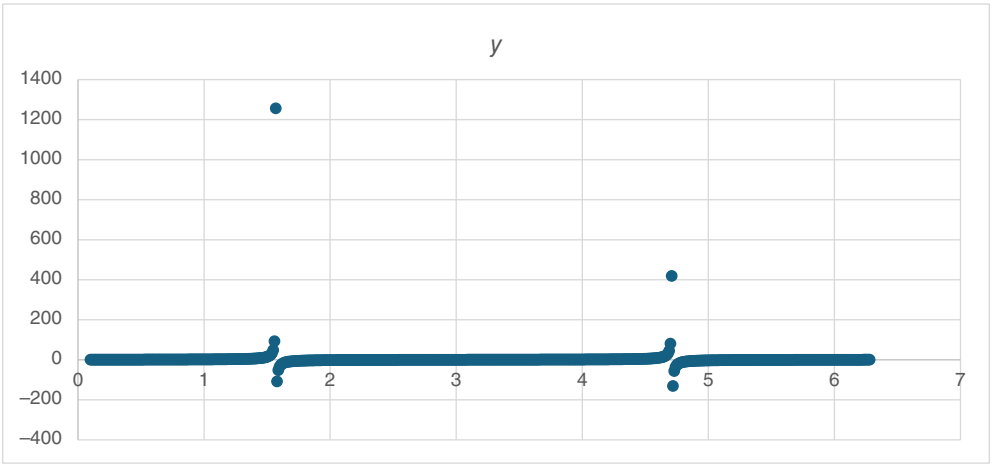
The resulting graph will be $y = \cos(x - \pi/3)$, with the maximum at $x = \pi/3$ and the minimums at $x = 4\pi/3$ and $7\pi/3$.

The complete sketch of the graph of $y = \cos(x - \pi/3)$ would look like this:

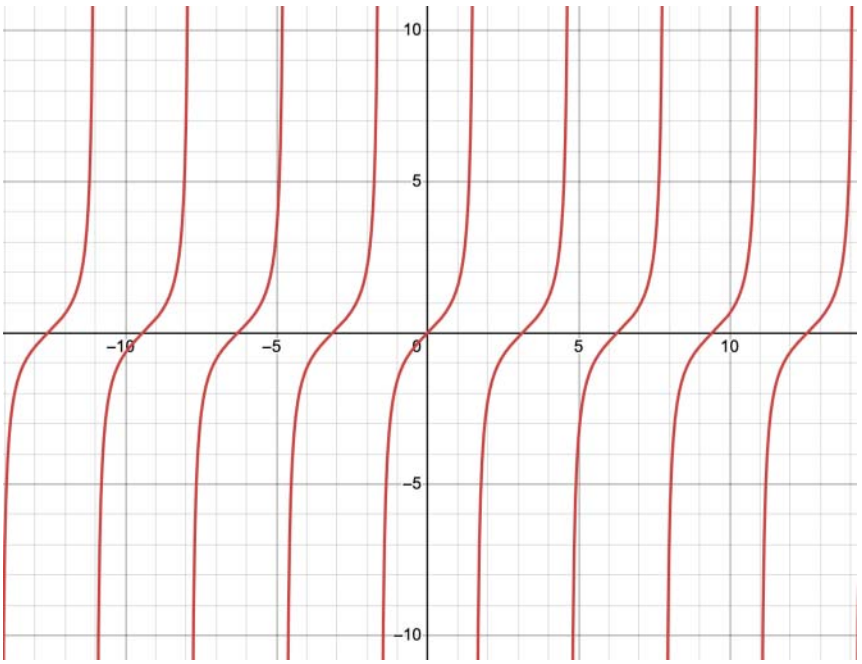


A similar process can be applied to the other trigonometric functions. We show the graph of the tangent function and leave the remaining function graphs to the exercises.

Repeating the process on $y = \tan(x)$, using values of x that begin at 0 and increase to 6.28 through increments of 0.01 radians, we find:



It is difficult, without additional computer assistance, to quite make out what is happening. However, at the values of $\pi/2$, $3\pi/2$, $5\pi/2$, and so on, the graph has "vertical asymptotes." The reason for this is that the function is undefined for those values of x.



7.2.8 The Inverse Trigonometric Functions

Recalling information from prerequisite courses, we recall functions have inverses if and only if they are one-to-one functions. Another way of describing this is that their graphs pass both the vertical and horizontal line tests (it is impossible to find a single vertical or horizontal line that passes through the graph at more than one point).

The trigonometric functions fail the horizontal line test in a rather spectacular way: they fail the test infinitely many times!

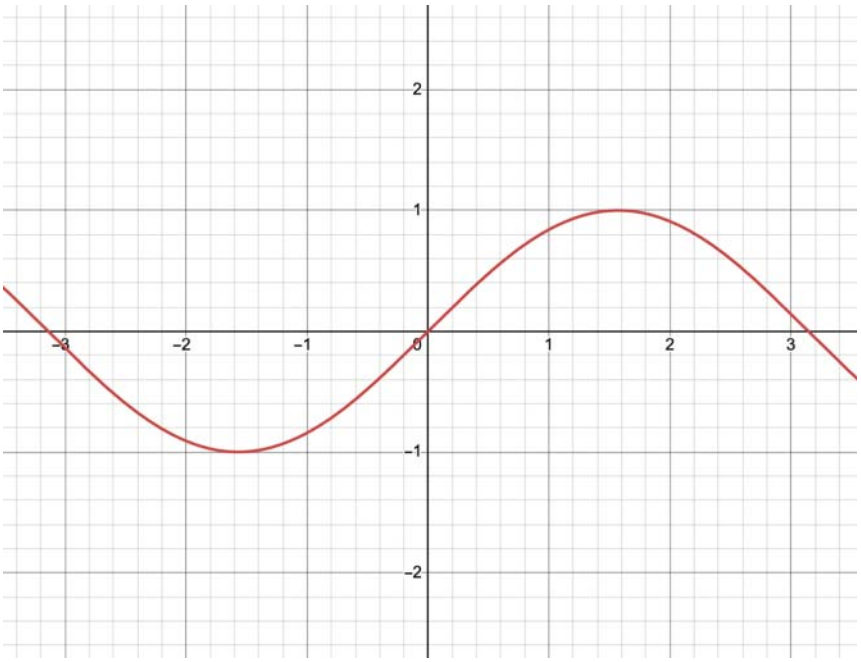
Nevertheless, we do speak of inverse trigonometric functions—how is this possible?

The answer lies in the concept of restricting the domain of the functions in a very specific manner, such that the restricted graph DOES pass the vertical and the horizontal line tests. This is not an obvious thing, but it is worth examination.

The notation for the "inverse sine function" takes one of two forms: $y = \arcsin(x)$ or $y = \sin^{-1} x$. The definition is very nuanced, as you will see.

We say $y = \arcsin(x)$ if y is the unique angle in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ such that $\sin y = x$.

Let us consider this for a moment and understand why that interval of values is chosen. Viewing the plane in terms of radian angles, the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ corresponds to the first and fourth quadrants of the plane. In the first quadrant, sine is positive. In the fourth quadrant, sine is negative (thus, we are deliberately choosing angles for which we can obtain the full extent of the values of the sine function). The pattern of sine values repeats itself infinitely often, and so we are isolating a single segment of the graph such that the horizontal and vertical line tests are not violated over that span.



Notice how the restricted portion of the sine graph takes on all possible values of sine between −1 and 1, inclusive, over this interval of x-values. Over this span, we define the inverse sine function.

Example 7.13 Find the exact value in radians of $\arcsin(-0.5)$.

Solution:

If we begin by saying $y = \arcsin(-0.5)$, we recognize that y is the unique angle in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ such that $\sin y = -0.5$.

The span of angles includes the first and fourth quadrants; since sine is positive in quadrant I and negative in quadrant IV, it follows that y is an angle in quadrant IV.

The value of 0.5 is a known result for the sine function; recall $\sin\frac{\pi}{6} = 0.5$, and therefore $\frac{\pi}{6}$ is the reference angle for the angle we seek. In quadrant IV, the angle with reference angle $\frac{\pi}{6}$ is the angle $-\frac{\pi}{6}$, and so $\arcsin(-0.5) = -\frac{\pi}{6}$.

Each of the inverse trigonometric functions is defined over a specific range of values, chosen in such a way that the trigonometric functions are invertible over those ranges. We present the inverses of the primary trigonometric functions here:

$y = \arcsin(x)$ if y is the unique angle in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ such that $\sin y = x$.

$y = \arccos(x)$ if y is the unique angle in the interval $[0, \pi]$ such that $\cos y = x$.

$y = \arctan(x)$ if y is the unique angle in the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ such that $\tan y = x$.

7.3 Introduction to Trigonometric Identities

A mathematical identity is an equation that holds true for any value of the variables involved in the equation. Trigonometric identities are equations involving trigonometric expressions that hold true for any angle. These identities describe the relationships between the trigonometric functions and can be used to simplify and manipulate trigonometric expressions. Understanding trigonometric identities is essential for solving more complex trigonometric problems and for converting complicated-looking trigonometric expressions into simpler ones.

7.3.1 Basic Trigonometric Identities

The most fundamental trigonometric identities are:

1) Pythagorean Identity:

$$\sin^2(\theta) + \cos^2(\theta) = 1$$

Reciprocal Identities:

$\sec(\theta) = 1/\cos(\theta)$

$\csc(\theta) = 1/\sin(\theta)$

$\cot(\theta) = 1/\tan(\theta)$

Cofunction Identities:

$\sin(\theta) = \cos(90° - \theta)$

$\cos(\theta) = \sin(90° - \theta)$

$\tan(\theta) = \cot(90° - \theta)$

Symmetric Identities:

$\sin(-\theta) = -\sin(\theta)$

$\cos(-\theta) = \cos(\theta)$

$\tan(-\theta) = -\tan(\theta)$

These identities can be derived directly from the definitions of the trigonometric functions and the properties of right triangles. We will establish several of the listed results and leave the proof of the others to the exercises.

Example 7.14 Prove the Pythagorean identity: $\sin^2(\theta) + \cos^2(\theta) = 1$

Proof:

The Pythagorean identity states that:

$$\sin^2(\theta) + \cos^2(\theta) = 1$$

To prove this, we can use the definitions of sine and cosine in a right triangle.

In a right triangle:

$$\sin(\theta) = \text{opposite/hypotenuse}$$
$$\cos(\theta) = \text{adjacent/hypotenuse}$$

Squaring both sides:

$$\sin^2(\theta) = (\text{opposite/hypotenuse})^2$$
$$\cos^2(\theta) = (\text{adjacent/hypotenuse})^2$$

Adding these together:

$$\sin^2(\theta) + \cos^2(\theta) = (\text{opposite/hypotenuse})^2 + (\text{adjacent/hypotenuse})^2$$

Now, we can use the Pythagorean theorem, which states that in a right triangle:

$$(\text{opposite})^2 + (\text{adjacent})^2 = (\text{hypotenuse})^2$$

Plugging this in:

$$\sin^2(\theta) + \cos^2(\theta) = (\text{opposite}^2/\text{hypotenuse}^2) + (\text{adjacent}^2/\text{hypotenuse}^2) = (\text{opposite}^2 + \text{adjacent}^2)/\text{hypotenuse}^2 = 1$$

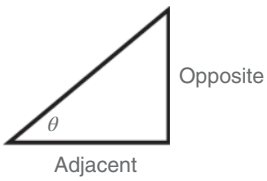
Therefore,

$$\sin^2(\theta) + \cos^2(\theta) = 1.$$

Example 7.15 Establish the identity $\sin(\theta) = \cos(90^\circ - \theta)$

Solution:

We draw a right triangle having acute angle θ:



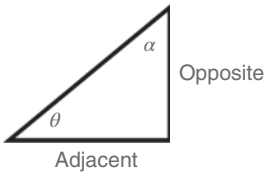
The three angles of any triangle sum to 180°, and the right angle is 90°. Therefore, if we refer to the unknown angle as α,

$$90^\circ + \theta + \alpha = 180^\circ$$

We can solve for α, finding:

$$\alpha = 90^\circ - \theta$$

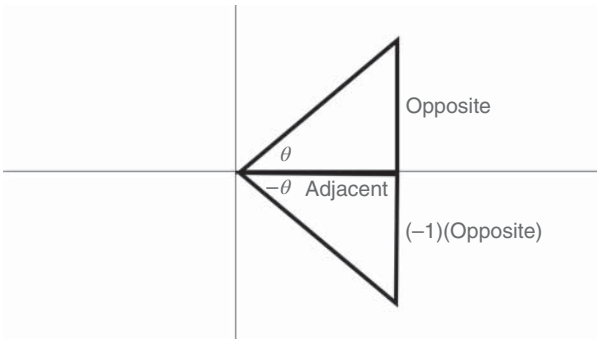
This establishes $\cos(90^\circ - \theta)$ is the same as $\cos\alpha$. From that perspective (and referring to the picture, observing the position of α):



Notice $\cos\alpha = \sin\theta$, the equality we were intending to establish.

Example 7.16 Establish $\tan(-\theta) = -\tan(\theta)$

Solution:
Again, we utilize a visual approach; one might suspect this technique is a bit "fishy" because of the specific placement of the right triangles in the plane, but the same method would apply no matter where the triangles were positioned.



Observe that for both triangles, the adjacent sides to the respective angles match precisely. The opposite sides have identical lengths, but the side for $-\theta$ has the opposite direction (and so its sign is reversed).

Notice, under this interpretation, $\tan(-\theta) = -\tan(\theta)$.

7.3.2 Using the Fundamental Identities in Computations

Example 7.17 Use a cofunction identity to find $\sin(60°)$.

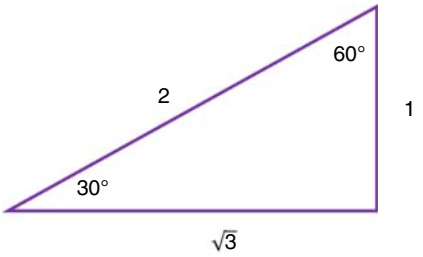
The cofunction identities for sine and cosine state that:

$$\sin(\theta) = \cos(90° - \theta)$$
$$\cos(\theta) = \sin(90° - \theta)$$

In this case, we want to find $\sin(60°)$; to do so, we can assign $\theta = 60°$ and use the cofunction identity:

$$\sin(60°) = \cos(90° - 60°) = \cos(30°)$$

Recalling the $1 - 2 - \sqrt{3}$ fundamental triangle:



We recall cos(30°) = adjacent/hypotenuse and so (for 30°)

adjacent = $\sqrt{3}$ units
opposite = 1 unit

Therefore:

cos(30°) = $\sqrt{3}/2$. Substituting into the cofunctions identity:

$$\sin(60°) = \cos(30°) = \sqrt{3}/2$$

So,

$$\sin(60°) = \sqrt{3}/2$$

Example 7.18 Use identities to simplify the expression: $\tan(x)/\sin(x)$

Solution:
First, let us recall the definitions of the trigonometric functions:

$\sin(x)$ = opposite/hypotenuse
$\cos(x)$ = adjacent/hypotenuse
$\tan(x)$ = opposite/adjacent

Now, let us plug these definitions into the expression:

$\tan(x)/\sin(x)$ = (opposite/adjacent)/(opposite/hypotenuse)
= (opposite/opposite) * (hypotenuse/adjacent)
= 1 * (hypotenuse/adjacent)
= hypotenuse/adjacent
= $1/\cos(x)$

Therefore, the simplified expression is:

$$\tan(x)/\sin(x) = 1/\cos(x)$$

This makes sense intuitively, as the tangent function is defined as the ratio of the opposite side to the adjacent side, while the sine function is defined as the ratio of the opposite side to the hypotenuse. Taking the ratio of these two gives the reciprocal of the cosine function.

Example 7.19 Use trigonometric identities to simplify the expression: $(\sin(x) + \cos(x))/$

$$(\sin(x) - \cos(x))(\sin(x) + \cos(x))/(\sin(x) - \cos(x))$$

Divide both the numerator and the denominator by $\cos(x)$:

$$(\sin(x)/\cos(x)+1)/(\sin(x)/\cos(x)-1)$$

Use the identity $\tan(x)=\sin(x)/\cos(x)$:

$$(\tan(x)+1)/(\tan(x)-1)$$

Factor the numerator and the denominator:

$$\begin{aligned}&(\tan(x)+1)/(\tan(x)-1)\\&=(\tan(x)+1)/(\tan(x)-1)*(\tan(x)+1)/(\tan(x)+1)\\&=(\tan^2(x)+2\tan(x)+1)/(\tan^2(x)-1)\end{aligned}$$

Use the Pythagorean identity $\sin^2(x)+\cos^2(x)=1$:

$$\begin{aligned}&(\tan^2(x)+2\tan(x)+1)/(\tan^2(x)-1)\\&=(1+2\tan(x)+1)/(1-1)\\&=2(1+\tan(x))\end{aligned}$$

Therefore, the simplified expression is:

$$(\sin(x)+\cos(x))/(\sin(x)-\cos(x))=2(1+\tan(x))$$

7.4 Trigonometric Functions of the Sum of Two Angles

In this section, we will explore the trigonometric functions of the sum of two angles. Understanding these identities is essential for solving a variety of trigonometric problems, including those involving compound angles, double angles, and half angles.

7.4.1 Sum and Difference Identities

The fundamental sum and difference identities for the trigonometric functions are as follows:

1) Sum Identities:
   - $\sin(A+B)=\sin(A)\cos(B)+\cos(A)\sin(B)$
   - $\cos(A+B)=\cos(A)\cos(B)-\sin(A)\sin(B)$
   - $\tan(A+B)=(\tan(A)+\tan(B))/(1-\tan(A)\tan(B))$
2) Difference Identities:
   - $\sin(A-B)=\sin(A)\cos(B)-\cos(A)\sin(B)$
   - $\cos(A-B)=\cos(A)\cos(B)+\sin(A)\sin(B)$
   - $\tan(A-B)=(\tan(A)-\tan(B))/(1+\tan(A)\tan(B))$

These identities can be derived using the definitions of the trigonometric functions and the properties of right triangles and will largely be left to the exercises. Consider, however, one such proof:

Example 7.20 Let us start with the geometric definition of sine:

$$\sin(A+B)=\text{adjacent/hypotenuse for the angle }(A+B)$$

We can express the adjacent and hypotenuse sides in terms of the individual angles A and B:

- adjacent $=\sin(A)\cos(B)+\cos(A)\sin(B)$
- hypotenuse $=\text{sqrt}(\sin^2(A)+\cos^2(A))*\text{sqrt}(\sin^2(B)+\cos^2(B))=1$

Substituting these into the geometric definition, we get:

$$\sin(A+B) = (\sin(A)\cos(B) + \cos(A)\sin(B))/1$$
$$\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$$

Therefore, the identity $\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$ is proven.

Second method

Start with the definition of the sine function:

$\sin(A+B) = y$-coordinate of the point on the unit circle at angle $(A+B)$.

Using the unit circle, we can express the point at angle $(A+B)$ in terms of the points at angles A and B:

$$x\text{-coordinate} = \cos(A)\cos(B) - \sin(A)\sin(B)$$
$$y\text{-coordinate} = \sin(A)\cos(B) + \cos(A)\sin(B)$$

Since $\sin(A+B)$ is the y-coordinate of the point at angle $(A+B)$, we have:

$$\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$$

Therefore, we have proven the identity:

$$\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B)$$

Example 7.21 Find $\sin 75^\circ = \sin(45^\circ + 30^\circ)$.

First, let us recall:

$$\sin(30^\circ) = 1/2$$
$$\cos(30^\circ) = \sqrt{3}/2$$
$$\sin(45^\circ) = \sqrt{2}/2$$
$$\cos(45^\circ) = \sqrt{2}/2$$

Therefore:

$$\begin{aligned}\sin(45^\circ + 30^\circ) &= \sin(45^\circ)\cos(30^\circ) + \cos(45^\circ)\sin(30^\circ)\\ &= (\sqrt{2}/2)(\sqrt{3}/2) + (\sqrt{2}/2)(1/2)\\ &= (\sqrt{6})/4 + \sqrt{2}/4\\ &= (\sqrt{6} + \sqrt{2})/4\end{aligned}$$

Simplifying further:

$$(\sqrt{6} + \sqrt{2})/4 = (2\sqrt{2} + \sqrt{2})/4 = 3\sqrt{2}/4$$

So:

$$\sin(45^\circ + 30^\circ) = 3\sqrt{2}/4.$$

Example 7.22 Find $\cos 15^\circ = \cos(60^\circ - 45^\circ)$.

First, let us recall

$$\sin(60^\circ) = \sqrt{3}/2$$
$$\cos(60^\circ) = 1/2$$
$$\sin(45^\circ) = \sqrt{2}/2$$
$$\cos(45^\circ) = \sqrt{2}/2$$

Therefore:

$$\cos(60^\circ - 45^\circ) = \cos(60^\circ)\cos(45^\circ) + \sin(60^\circ)\sin(45^\circ)$$
$$= (1/2)(\sqrt{2}/2) + (\sqrt{3}/2)(\sqrt{2}/2)$$
$$= \sqrt{2}/4 + \sqrt{6}/4$$
$$= (\sqrt{2} + \sqrt{6})/4$$

Simplifying further:

$$(\sqrt{2} + \sqrt{6})/4 = (2\sqrt{2} + 2\sqrt{3})/8 = \sqrt{2} + \sqrt{3}$$

So:

$$\cos(60^\circ - 45^\circ) = \sqrt{2} + \sqrt{3}.$$

7.5 Trigonometric Functions of the Difference of Two Angles

In the previous section, we explored the sum identities for trigonometric functions. In this section, we will focus on the difference identities, which are equally important in solving various trigonometric problems.

7.5.1 Difference Identities

The fundamental difference identities for the trigonometric functions are as follows:

Difference Identities:

- $\sin(A - B) = \sin(A)\cos(B) - \cos(A)\sin(B)$
- $\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B)$
- $\tan(A - B) = (\tan(A) - \tan(B))/(1 + \tan(A)\tan(B))$

These identities can be derived using the definitions of the trigonometric functions and the properties of right triangles, and the proofs are left to the exercises.

Example 7.23 Find $\sin(60^\circ - 30^\circ)$, using the difference of angles identities.

We know the following identities:

$$\sin(A + B) = \sin(A)\cos(B) + \cos(A)\sin(B)$$
$$\sin(A - B) = \sin(A)\cos(B) - \cos(A)\sin(B)$$

Let's plug in $A = 60^\circ$ and $B = 30^\circ$ into the $\sin(A - B)$ formula:

$$\sin(60^\circ - 30^\circ) = \sin(60^\circ)\cos(30^\circ) - \cos(60^\circ)\sin(30^\circ)$$

Now, let us find the values of $\sin(60^\circ)$, $\cos(30^\circ)$, $\cos(60^\circ)$, and $\sin(30^\circ)$:

$$\sin(60^\circ) = \sqrt{3}/2$$
$$\cos(30^\circ) = \sqrt{3}/2$$
$$\cos(60^\circ) = 1/2$$
$$\sin(30^\circ) = 1/2$$

Plugging these values into the formula:

$$\begin{aligned}\sin(60^\circ - 30^\circ) &= (\sqrt{3}/2)(\sqrt{3}/2) - (1/2)(1/2)\\ &= (3/4) - (1/4)\\ &= 1/2\end{aligned}$$

Therefore, the final result is:

$$\sin(60^\circ - 30^\circ) = 1/2.$$

Example 7.24 Find $\cos(45^\circ - 30^\circ)$.

We know the following identities:

$$\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$
$$\cos(A - B) = \cos(A)\cos(B) + \sin(A)\sin(B)$$

Let's plug in $A = 45^\circ$ and $B = 30^\circ$ into the $\cos(A - B)$ formula:

$$\cos(45^\circ - 30^\circ) = \cos(45^\circ)\cos(30^\circ) + \sin(45^\circ)\sin(30^\circ)$$

Now, let us find the values of $\cos(45^\circ)$, $\cos(30^\circ)$, $\sin(45^\circ)$, and $\sin(30^\circ)$:

$$\cos(45^\circ) = \sqrt{2}/2$$
$$\cos(30^\circ) = \sqrt{3}/2$$
$$\sin(45^\circ) = \sqrt{2}/2$$
$$\sin(30^\circ) = 1/2$$

Plugging these values into the formula:

$$\begin{aligned}\cos(45^\circ - 30^\circ) &= (\sqrt{2}/2)(\sqrt{3}/2) + (\sqrt{2}/2)(1/2)\\ &= (\sqrt{6} + \sqrt{2})/4\\ &= (\sqrt{6} + \sqrt{2})/2\sqrt{2}\end{aligned}$$

Therefore, the final result is:

$$\cos(45^\circ - 30^\circ) = (\sqrt{6} + \sqrt{2})/2\sqrt{2}.$$

Example 7.25 Using the identities to evaluate $\sin(60^\circ - 30^\circ)$ and $\cos(60^\circ - 30^\circ)$

Recall:

$$\sin(60^\circ) = \sqrt{3}/2$$
$$\cos(60^\circ) = 1/2$$
$$\cos(30^\circ) = \sqrt{3}/2$$
$$\sin(30^\circ) = 1/2$$

Plugging these values into the difference of angles formula for sine:

$$\begin{aligned}\sin(60^\circ - 30^\circ) &= (\sqrt{3}/2)(\sqrt{3}/2) - (1/2)(1/2)\\ &= (3/4) - (1/4)\\ &= 1/2\end{aligned}$$

Therefore, $\sin(60^\circ - 30^\circ) = 1/2$.

$\cos(60^\circ - 30^\circ)$:

$\cos(60^\circ - 30^\circ) = \cos(60^\circ)\cos(30^\circ) + \sin(60^\circ)\sin(30^\circ)$

Using the same trigonometric values as before:

$$\begin{aligned}\cos(60° - 30°) &= (1/2)(\sqrt{3}/2) + (\sqrt{3}/2)(1/2) \\ &= (\sqrt{3}/4) + (\sqrt{3}/4) \\ &= \sqrt{3}/2\end{aligned}$$

Therefore, $\cos(60° - 30°) = \sqrt{3}/2$.

7.6 Trigonometric Functions of Double-Angles

In the previous section, we explored the difference identities for trigonometric functions. In this section, we will focus on the double-angle identities, which are useful in a variety of trigonometric applications.

7.6.1 Double-Angle Identities

The fundamental double-angle identities for the trigonometric functions are as follows:

Double-Angle Identities:

- $\sin(2A) = 2\sin(A)\cos(A)$
- $\cos(2A) = \cos^2(A) - \sin^2(A) = 2\cos^2(A) - 1 = 1 - 2\sin^2(A)$
- $\tan(2A) = 2\tan(A)/(1 - \tan^2(A))$

These identities can be derived using the definitions of the trigonometric functions and the properties of right triangles and are (with the exception of the double angle for sine) left to the exercises.

Example 7.26 Show $\sin(2A) = 2\sin(A)\cos(A)$.

Solution:
We can start with the definition of the sine of a double angle:

$\sin(2A) = 2\sin(A)\cos(A)$

To prove this, we can use the trigonometric identity for the sine of the sum of two angles:

$\sin(A + B) = \sin(A)\cos(B) + \cos(A)\sin(B)$

Letting $A = A$ and $B = A$, we get:

$\sin(A + A) = \sin(A)\cos(A) + \cos(A)\sin(A)$

Simplifying the left-hand side, we get:

$\sin(2A) = 2\sin(A)\cos(A)$

Therefore, we have proven the identity:

$\sin(2A) = 2\sin(A)\cos(A)$.

Example 7.27 Find $\sin(2\theta)$ if $\sin(\theta) = 3/5$ and θ lies in the first quadrant.

To do this, we can use the trigonometric identity for the sine of double an angle:

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta)$$

We know that $\sin(\theta) = 3/5$, so we need to find $\cos(\theta)$ as well.
Using the Pythagorean identity:

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

Plugging in $\sin(\theta) = 3/5$, we get:

$$\cos^2(\theta) + (3/5)^2 = 1$$
$$\cos^2(\theta) + 9/25 = 1$$
$$\cos^2(\theta) = 16/25$$
$$\cos(\theta) = \pm 4/5$$

Since sine is a positive value, we know that the angle θ is in the first or second quadrant, so the positive value of $\cos(\theta)$ is $4/5$.
Now, we can plug the values of $\sin(\theta)$ and $\cos(\theta)$ into the identity for $\sin(2\theta)$:

$$\begin{aligned}\sin(2\theta) &= 2\sin(\theta)\cos(\theta)\\ &= 2(3/5)(4/5)\\ &= 24/25\end{aligned}$$

Therefore, if $\sin(\theta) = 3/5$, then $\sin(2\theta) = 24/25$.

Example 7.28 Find $\cos(2\theta)$ if $\cos(\theta) = 4/5$ and θ lies in quadrant IV.
To do this, we can use the trigonometric identity for the cosine of double an angle:

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta)$$

We know that $\cos(\theta) = 4/5$, so we need to find $\sin(\theta)$ as well.
Using the Pythagorean identity:

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

Plugging in $\cos(\theta) = 4/5$, we get:

$$(4/5)^2 + \sin^2(\theta) = 1$$
$$16/25 + \sin^2(\theta) = 1$$
$$\sin^2(\theta) = 9/25$$
$$\sin(\theta) = \pm 3/5$$

Since angle θ is in the fourth quadrant, the value of $\sin(\theta)$ is $-3/5$.
Now, we can plug the values of $\cos(\theta)$ and $\sin(\theta)$ into the identity for $\cos(2\theta)$:

$$\begin{aligned}\cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta)\\ &= (4/5)^2 - (-3/5)^2\\ &= 16/25 - 9/25\\ &= 7/25\end{aligned}$$

Therefore, if $\cos(\theta) = 4/5$, then $\cos(2\theta) = 7/25$.

7.7 Trigonometric Functions of Half-Angles

7.7.1 Half-Angle Identities

The fundamental half-angle identities for the trigonometric functions are as follows:
Half-Angle Identities:

- $\sin(A/2) = \pm\sqrt{((1-\cos(A)) / 2)}$
- $\cos(A/2) = \pm\sqrt{((1+\cos(A)) / 2)}$
- $\tan(A/2) = \pm\sqrt{((1-\cos(A))/(1+\cos(A)))}$

These identities can be derived using the definitions of the trigonometric functions and the properties of right triangles and are largely left to the exercises.

Example 7.29 Proving the identity for $\sin(A/2)$.
We want to prove the identity:

$$\sin(A/2) = \pm\sqrt{[(1-\cos(A))/2]}$$

Proof:
Start with the identity:

$$\sin(A/2) = \pm\sqrt{[(1-\cos(A)) / 2]}$$

Square both sides to get:

$$\sin^2(A/2) = (1-\cos(A)) / 2$$

Now, we can use the trigonometric identity:

$$\sin^2(A/2) = (1-\cos(A)) / 2$$

Rearranging the terms, we get:

$$2\sin^2(A/2) = 1-\cos(A)$$

Taking the square root of both sides, we get:

$$\pm\sqrt{(2)}\sin(A/2) = \pm\sqrt{(1-\cos(A))}$$

Dividing both sides by $\pm\sqrt{2}$, we get:

$$\sin(A/2) = \pm\sqrt{[(1-\cos(A)) / 2]}$$

Therefore, we have proven the identity:

$$\sin(A/2) = \pm\sqrt{[(1-\cos(A)) / 2]}$$

Example 7.30 Proving the identity for $\cos(A/2)$
We want to prove the identity:

$$\cos(A/2) = \pm\sqrt{[(1+\cos(A)) / 2]}$$

Proof:
Start with the Pythagorean identity:

$$\sin^2(A/2) + \cos^2(A/2) = 1$$

Rearranging the terms, we get:

$$\cos^2(A/2) = 1 - \sin^2(A/2)$$

Substituting the expression for $\sin(A/2)$ from the previous identity:

$$\cos^2(A/2) = 1 - (1 - \cos(A)) / 2$$
$$\cos^2(A/2) = (1 + \cos(A)) / 2$$

Taking the square root of both sides, we get:

$$\cos(A/2) = \pm\sqrt{[(1 + \cos(A)) / 2]}$$

Therefore, we have proven the identity:

$$\cos(A/2) = \pm\sqrt{[(1 + \cos(A)) / 2]}$$

Example 7.31 Find $\sin(\pi/6)$.

On the unit circle, the angle of $\pi/6$ radians (or 30 degrees) corresponds to the following coordinates:

$$x = \cos(\pi/6)$$
$$y = \sin(\pi/6)$$

We can use the Pythagorean identity to find $\sin(\pi/6)$:

$$\sin^2(\pi/6) + \cos^2(\pi/6) = 1$$

Substituting the values, we get:

$$\sin^2(\pi/6) + (\sqrt{3}/2)^2 = 1$$
$$\sin^2(\pi/6) + 3/4 = 1$$
$$\sin^2(\pi/6) = 1/4$$
$$\sin(\pi/6) = \sqrt{(1/4)} = \sqrt{(1/2)} / \sqrt{2} = 1/\sqrt{3}$$

Therefore, the value of $\sin(\pi/6)$ is $1/\sqrt{3}$.

We can also express this as an approximate decimal value:

$$\sin(\pi/6) \approx 0.5773$$

So, the final answer is:

$$\sin(\pi/6) = 1/\sqrt{3} \approx 0.5773.$$

Example 7.32 Find $\cos(\pi/4)$.

On the unit circle, the angle of $\pi/4$ radians (or 45 degrees) corresponds to the following coordinates:

$$x = \cos(\pi/4)$$
$$y = \sin(\pi/4)$$

We can use the Pythagorean identity to find $\cos(\pi/4)$:

$$\sin^2(\pi/4) + \cos^2(\pi/4) = 1$$

Substituting the values, we get:

$$(\sqrt{2}/2)^2 + \cos^2(\pi/4) = 1$$
$$1/2 + \cos^2(\pi/4) = 1$$
$$\cos^2(\pi/4) = 1/2$$
$$\cos(\pi/4) = \sqrt{(1/2)} = \sqrt{2}/2$$

Therefore, the value of $\cos(\pi/4)$ is $\sqrt{2}/2$.

We can also express this as an approximate decimal value:

$$\cos(\pi/4) \approx 0.7071$$

So, the final answer is:

$$\cos(\pi/4) = \sqrt{2}/2 \approx 0.7071$$

Example: Using the identities to evaluate $\sin(30°)$ and $\cos(30°)$

$$\sin(30°) = \sin(\pi/6)$$
$$\sin(A/2) = \pm\sqrt{[(1 - \cos(A))/2]}$$
$$\sin(\pi/6) = \pm\sqrt{[(1 - \cos(\pi/6))/2]}$$
$$\sin(\pi/6) = \pm\sqrt{[(1 - \sqrt{3}/2)/2]}$$
$$\sin(\pi/6) = \pm\sqrt{(1/4)}$$
$$\sin(\pi/6) = \pm 1/\sqrt{2}$$
$$\sin(30°) = 1/\sqrt{2} \approx 0.5$$
$$\cos(30°) = \cos(\pi/6)$$
$$\cos(A/2) = \pm\sqrt{[(1 + \cos(A))/2]}$$
$$\cos(\pi/6) = \pm\sqrt{[(1 + \cos(\pi/6))/2]}$$
$$\cos(\pi/6) = \pm\sqrt{[(1 + \sqrt{3}/2)/2]}$$
$$\cos(\pi/6) = \pm\sqrt{(3/4)}$$
$$\cos(\pi/6) = \pm\sqrt{3}/2$$
$$\cos(30°) = \sqrt{3}/2 \approx 0.8660$$

Therefore, the final answers are:

$$\sin(30°) = 1/\sqrt{2} \approx 0.5$$
$$\cos(30°) = \sqrt{3}/2 \approx 0.8660.$$

Example 7.33 Given: $\cos(A) = 3/5$, where A is an angle lying in quadrant IV, use the half-angle identities to find the trigonometric values $\sin(A/2)$, $\cos(A/2)$, and $\tan(A/2)$ given that $\cos(A) = 3/5$.

Solution:

Observe the specification that angle A lies in quadrant IV tells us $270° < A < 360°$ and thus $135° < \frac{A}{2} < 180°$. This is important because it tells us $\frac{A}{2}$ lies in quadrant II, which will prove essential in what follows.

Next, we employ the half-angle formula:

$$\sin(A/2) = \pm\sqrt{((1 - \cos(A))/2)}$$

$$\sin\left(\frac{A}{2}\right) = \pm\sqrt{\frac{1 - \frac{3}{5}}{2}} = \pm\sqrt{\frac{2}{10}} = \pm\sqrt{\frac{1}{5}}$$

Since $A/2$ lies in the second quadrant (where sine is positive), we know we want $\sqrt{\frac{1}{5}}$.

NOTE: This type of reasoning is rather elaborate; be sure you follow the logic for each of the steps!

Similar reasoning can be followed to determine the cosine of $A/2$, but we have an alternative, the Pythagorean identity.

$$\left(\sqrt{\frac{1}{5}}\right)^2 + \cos^2\frac{A}{2} = 1$$
$$\frac{1}{5} + \cos^2\frac{A}{2} = 1$$
$$\cos^2\frac{A}{2} = \frac{4}{5}$$

We will use the square root property, keeping in mind $A/2$ is in the second quadrant, and so we will use the negative alternative:

$$\cos\frac{A}{2} = -\sqrt{\frac{4}{5}}$$
$$\cos\frac{A}{2} = -\frac{2\sqrt{5}}{5}$$

Finally, using the identity $\tan\theta = \frac{\sin\theta}{\cos\theta}$, we have

$$\tan\frac{A}{2} = \frac{\sqrt{\frac{1}{5}}}{-\sqrt{\frac{4}{5}}}$$
$$\tan\frac{A}{2} = -\sqrt{\frac{1}{4}}$$
$$\tan\frac{A}{2} = -\frac{1}{2}$$

Example 7.34 Given $\sin(A) = 3/5$, use the half-angle identities to find the trigonometric values $\sin(A/2)$,

$\cos(A/2)$, and $\tan(A/2)$ given that $\sin(A) = 3/5$.

Solution:

Step 1: Find the value of A.

From the given information, we have $\sin(A) = 3/5$.

Using the inverse sine function, we can find the value of A:

$$A = \sin^{-1}(3/5) = 53.13^\circ$$

Step 2: Find $\sin(A/2)$ and $\cos(A/2)$ using the half-angle identities.

$$\sin(A/2) = \pm\sqrt{[(1 - \cos(A)) / 2]}$$
$$\sin(A/2) = \pm\sqrt{[(1 - 3/5) / 2]}$$
$$\sin(A/2) = \pm\sqrt{(2/5)}$$
$$\sin(A/2) = \pm\sqrt{(2)}/\sqrt{5}$$
$$\cos(A/2) = \pm\sqrt{[(1 + \cos(A)) / 2]}$$
$$\cos(A/2) = \pm\sqrt{[(1 + 3/5) / 2]}$$
$$\cos(A/2) = \pm\sqrt{(4/5)}$$
$$\cos(A/2) = \pm 2/\sqrt{5}$$

Step 3: Find $\tan(A/2)$ using the identities.

$$\tan(A/2) = \sin(A/2)/\cos(A/2)$$
$$\tan(A/2) = \pm\sqrt{(2)}/\sqrt{5}/\pm 2/\sqrt{5}$$
$$\tan(A/2) = \pm 1/\sqrt{2}$$

Therefore, the final values are:

$$\sin(A/2) = \pm\sqrt{(2)}/\sqrt{5}$$
$$\cos(A/2) = \pm 2/\sqrt{5}$$
$$\tan(A/2) = \pm 1/\sqrt{2}.$$

8

Introduction to Complex Numbers

8.1 Introduction to Complex Numbers

Real numbers (the vast collection of natural numbers, whole numbers, integers, rational numbers, and irrational numbers) have their limitations when it comes to the solutions to some types of equations. As an illustration, consider:

$$x^2 + 3x + 11 = 0$$

If we attempt to solve this equation, we find it is impossible to solve by the method of factoring, and so we turn to the quadratic formula. In this case, $a = 1$, $b = 3$, $c = 11$, so the formula yields:

$$x = \frac{-3 \pm \sqrt{3^2 - 4(1)(11)}}{2(1)}$$

$$x = \frac{-3 \pm \sqrt{-35}}{2}$$

Since the radicand of the square root is negative, it is impossible to simplify this result further by using any of the methods seen thus far. Nevertheless, it becomes important in applications for us to be able to establish a solution to the equation, and for this reason, a new set of values, the complex numbers, is introduced.

In other words, complex numbers are motivated by the need to simplify square (and higher even-index) roots of negative numbers. In mathematics, it is often necessary to introduce a new concept to resolve a particular problem; doing so spawns consequences unforeseeable when the definition is initially made, which (in this case) leads to a rich field of mathematics called Complex Analysis.

Complex numbers are an important extension of the real number system, allowing for the representation and manipulation of quantities that cannot be fully described or expressions that cannot be simplified using only real numbers. In this section, we will explore the fundamental properties and applications of complex numbers.

One might speculate this concept is limited to abstract solutions of equations, but this is not true. There are countless applications to real-world situations that require complex numbers for their study. An illustration of the use of complex numbers is the two-dimensional steady-state heat transfer problem; the solution of this problem uses methods we have seen, as well as additional techniques and concepts that lie beyond the scope of our presentation here.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

8.1.1 The Problem to Be Solved

Referring to our example of the quadratic equation given earlier, what is the square root of a negative number, such as $\sqrt{-35}$?

What is the fourth root of a negative number, such as $\sqrt[4]{-65.2}$?

We know that, by definition, the values described here are the numbers x and y such that $x^2 = -35$ and that $y^4 = -65.2$. Unfortunately, an even power of any real number yields a positive result. How are we to overcome the seeming impossibility of finding a solution to such problems?

To this point in our studies, we would simply terminate our computations and proclaim there is "no solution" to the problem. However, as has been mentioned, solving particular application problems requires us to find a solution to these equations, and this leads to the topic of this chapter.

8.1.2 What Are Complex Numbers?

A complex number is a number that can be expressed in the form $a + bi$, where a and b are real numbers, and i is the imaginary unit, where $\boldsymbol{i}$ is defined as that number such that $i^2 = -1$; an equivalent definition is that $i = \sqrt{-1}$. The real number a is called the real part of the complex number, and the real number b is called the imaginary part.

NOTE: In the current sense, the word "unit" refers to a measure of 1. Therefore, the imaginary unit is describing the concept of "1" as it applies to complex numbers.

Some examples of complex numbers are:

- $3 + 2i$,
- $-\sqrt{5} + 3i$,
- $-5 + 0i$ (which is the same as -5),
- $0 + 4i$ (which is the same as $4i$).

The last two examples present situations that call for special terminology. When the imaginary part of a complex number is zero, we refer to the complex number as "real." When the real part of a complex number is zero, we refer to the complex number as "pure imaginary."

Introduction of this concept allows us to simplify square roots of negative numbers; higher-index roots of negative numbers will not be addressed by us here, as that challenge is better met in a course on complex variables.

Example 8.1 Simplify $\sqrt{-35}$, using complex numbers.

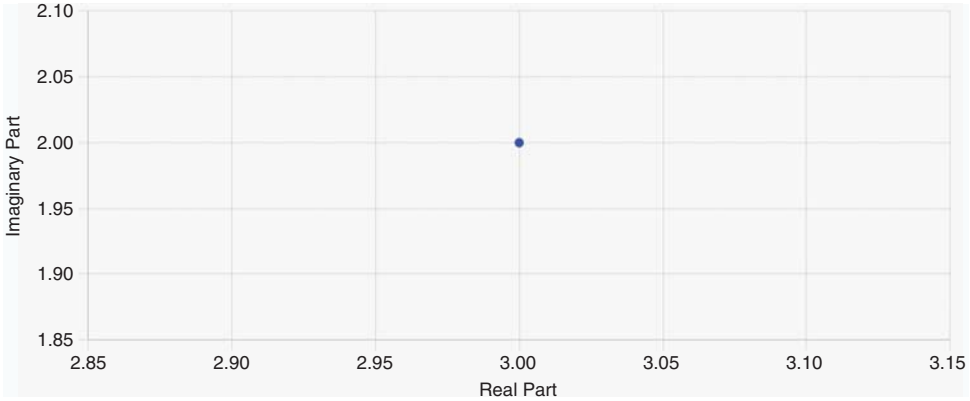
Solution: Observe $\sqrt{-35}$ can be expressed as $\sqrt{(-1)(35)}$, which (in turn) can be separated (using the product rule of radicals) to be $\sqrt{-1} \cdot \sqrt{35}$. By incorporating the definition of the imaginary unit, $i = \sqrt{-1}$, we can simplify to $i\sqrt{35}$. The remaining square root cannot be simplified (the prime factorization of 35 is $5 \cdot 7$, and so no perfect squares are factors of 35), and thus $\sqrt{-35} = i\sqrt{35}$.

While it is acceptable to write $\sqrt{-35} = \sqrt{35 \cdot (-1)} = \sqrt{-35} = \sqrt{35}\sqrt{-1} = \sqrt{-35} = \sqrt{35i}$, care must be taken to ensure i is not carelessly placed within the radicand; doing so introduces the value of $\sqrt{35i}$, which requires knowledge of complex variables to simplify (and is NOT equivalent to $i\sqrt{35}$). The student is cautioned to ensure the imaginary unit i is NOT placed within the radicand.

8.1.3 The Complex Plane

Complex numbers can be visualized on a special form of a coordinate plane, known as the **complex plane** or as an **Argand diagram**. In this representation, the real part is plotted on the horizontal

axis, and the imaginary part is plotted on the vertical axis. For example, the complex number $3 + 2i$ would be represented by the point (3, 2) on the complex plane.



Aside: The alternative name for the complex plane, an Argand diagram, refers to the mathematician Jean-Robert Argand (1768–1822). This is an example of an individual who "got the credit" for an idea another person introduced earlier in history. In the case of the complex plane, such a construction was previously proposed by Caspar Wessel (1745–1818). While this is not a terrifically important distinction to make, it does point out how history (even in such a regimented and detailed field such as mathematics) can take unusual turns.

8.2 Addition and Subtraction of Complex Numbers

8.2.1 Addition of Complex Numbers

To add two complex numbers $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$, we add the corresponding real and imaginary parts:

$$z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i.$$

Example 8.2 Find the sum of the complex numbers $(3 + 2i)$ and $(4 - i)$.
To add complex numbers, we add the real parts and the imaginary parts separately.
Real part: $3 + 4 = 7$
Imaginary part: $2i + (-i) = 2 - i = i$
Therefore, the sum of the two complex numbers is $7 + i$.

Example 8.3 Add the complex numbers: $(-2 + 3i)$ and $(5 - 4i)$.
Real part: $-2 + 5 = 3$
Imaginary part: $3i - 4i = -i$
Therefore, the sum of the two complex numbers is $3 - i$.

At some point, "elementary" computations become tedious to express in fully written-out form. While we should not omit steps the outcome of which we are absolutely certain, it becomes expedient to limit our explicit presentation of all steps. The subdivision of the problem into real and imaginary parts can (and should) be done using mental arithmetic, and thus can be suppressed when performing the computation.

Example 8.4 Find the sum of the complex numbers $z_1 = 3 + 2i$ and $z_2 = -1 + 4i$:

$$z_1 + z_2 = (3 + (-1)) + (2 + 4)i = 2 + 6i$$

Example 8.5 Find the sum of the complex numbers $(3 + 2i)$ and $(4 - i)$.

$$\begin{aligned}(3 + 2i) + (4 - i) &= (3 + 4) + (2i - i)\\ &= 7 + i\end{aligned}$$

It is possible in the addition of complex numbers that one or both of the parts of the complex number will be annihilated. This is a technical term for the more casual phrase "subtracted out." Consider, for example, $(5 + 2i) + (3 - 2i) = 8 + 0i = 8$. When this addition is performed, the imaginary parts are "annihilated." While we can express the sum simply as 8, it is customary when working with complex numbers to insist the solution be written in $a + bi$ form, which highlights the fact that we are not simply working with real numbers.

8.2.2 Subtraction of Complex Numbers

To subtract two complex numbers $z_1 = a_1 + b_1 i$ and $z_2 = a_2 + b_2 i$, we subtract the corresponding real and imaginary parts: $z_1 - z_2 = (a_1 - a_2) + (b_1 - b_2)i$. Particular care should be taken when the imaginary part of the subtrahend (the value being subtracted) is negative, as the distributive law will change the sign of that part of the second complex number.

For example, let us subtract the complex numbers $z_1 = 3 + 2i$ and $z_2 = -1 + 4i$

$$z_1 - z_2 = (3 - (-1)) + (2 - 4)i = 4 - 2i$$

Example 8.6 Subtract the complex number $(5 - 3i)$ from $(8 + 2i)$.

$$\begin{aligned}(8 + 2i) - (5 - 3i) &= (8 - 5) + (2i + 3i)\\ &= 3 + 5i\end{aligned}$$

So, the difference between $(8 + 2i)$ and $(5 - 3i)$ is $3 + 5i$.

It is worth mentioning that this subdivision into real and imaginary parts is not, in practice, needed. If we were to regard the complex numbers as being similar to linear expressions, we can calculate the difference by combining like terms.

That is, for $(8 + 2i) - (5 - 3i)$, we can simplify first to $8 + 2i - 5 + 3i$ and then combine the like terms as $8 - 5 + 2i - 3i = 3 - 1i = 3 - i$. This is viewed by some as being more readily understood, though ultimately it is a distinction without a difference: either method will produce the same result, and we will employ one or the other approaches as seems appropriate for a given problem.

Example 8.7 Subtract the complex numbers: $(5 + 2i) - (3 - i)$.

Real part: $5 - 3 = 2$

Imaginary part: $2i - (-i) = 3i$

Therefore, the difference of the two complex numbers is $2 + 3i$.

Example 8.8 Subtract the complex numbers: $(4 - 3i) - (2 + i)$.

Real part: $4 - 2 = 2$

Imaginary part: $-3i - i = -4i$

Therefore, the difference of the two complex numbers is $2 - 4i$.

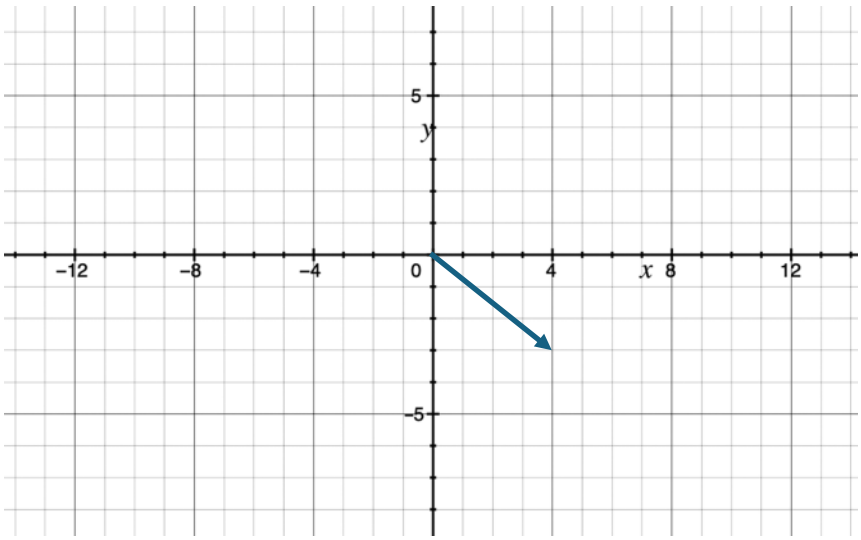
In both addition and subtraction, the key observation we must make is to separate the real and imaginary parts, perform the operations, and then combine the results back into a single complex number of the form a + bi.

8.2.3 Geometric Interpretation of Complex Numbers as Points or Vectors

As mentioned, we can represent complex numbers on the complex plane, where the real part is the x-coordinate, and the imaginary part is the y-coordinate. Addition and subtraction of complex numbers can be interpreted geometrically on the complex plane, as we will now show. The utility of doing so is not immediately apparent, but will be seen to offer tremendous insight to physical applications of the concept.

The addition is often presented geometrically as a sum of vectors. A **vector** is a directed line segment (that is, a line segment for which one endpoint is identified as the initial point and the other endpoint as the terminal point), drawn in such a way that an arrowhead indicates the terminal point of the segment. If a vector is positioned in such a way that the initial point should lie at the origin (the point (0,0)), then the vector is said to be in **standard position**.

In this construction, the initial point is sometimes referred to as the "tip" and the terminal point as the "tail." As an example, the complex number 4 − 3i can be represented visually by the following vector:



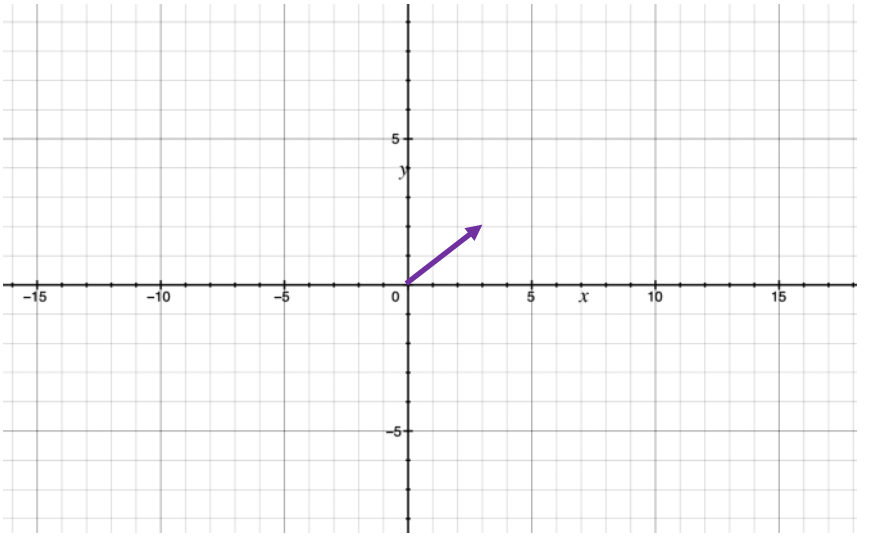
Observe the initial point (the tip) is positioned at the origin, and the terminal point (the tail) is at the point (4, −3), which refers to the real and imaginary parts of 4 − 3i.

The sum of two complex numbers is the terminal point of the vector sum of the two complex numbers on the complex plane. By the term "vector sum," we are referring to a new vector created using the "tip to tail addition method of vectors." Observe how the earlier references to "tip" and "tail" becomes significant when performing the addition we are about to undertake,

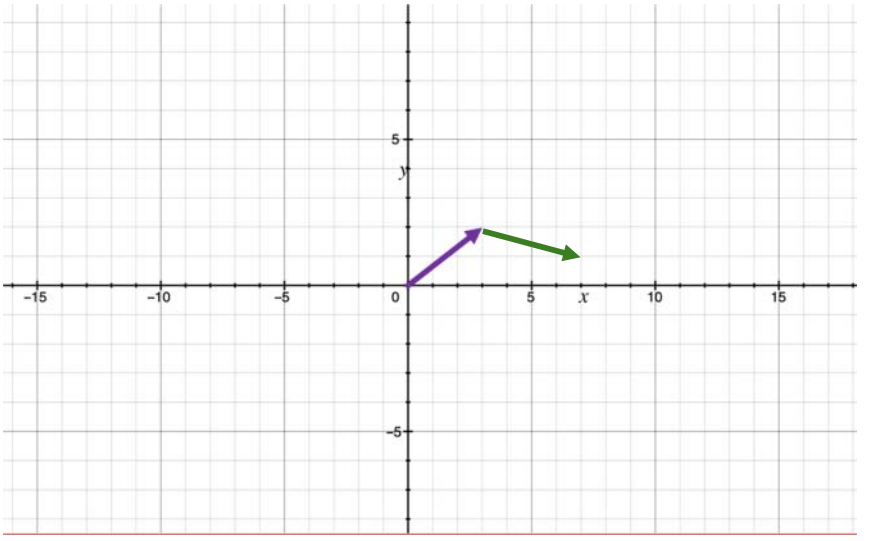
An example will illustrate the method far better than a written description; let us examine the problem given earlier:

Example 8.9 $(3 + 2i) + (4 - i)$.

First, set up the complex plane and draw the first vector given in the problem.



The second complex number $(4 - i)$ will be overlaid upon this complex plane, positioned in such a way that its tip aligns with the tail of the first complex number.

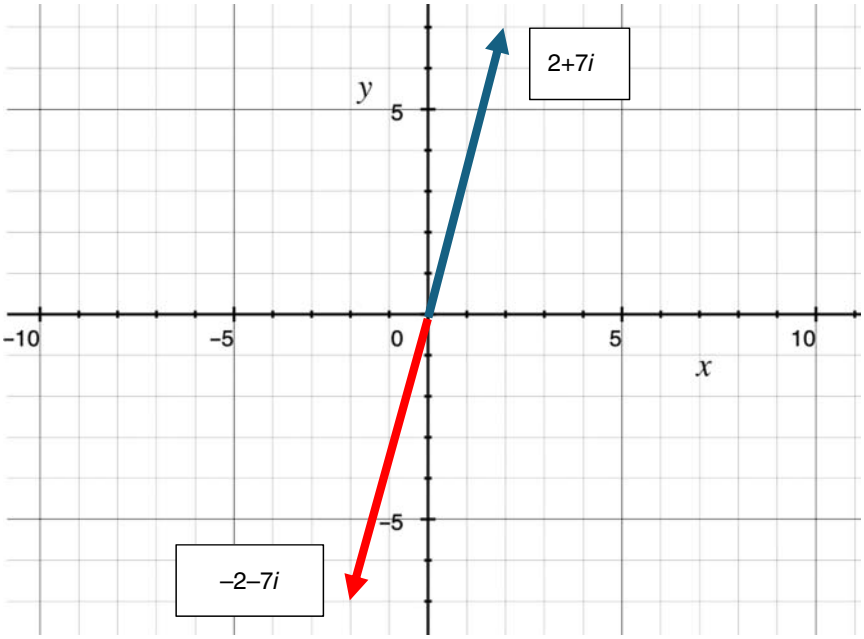


Note the location of the terminal point of the combined vectors: that point has coordinates (7,1), which corresponds to the complex number $7 + i$, the sum calculated earlier.

For subtraction, the difference between two complex numbers is the point that is the vector difference of the two complex numbers on the complex plane. In order to understand the visualization, we must first consider the "negation" of a vector, and to recall that subtraction of the form A − B is equivalent to A + (−B).

The negation of a vector reverses the direction of that vector by 180°, or π radians. To see this, we consider the vector $2 + 7i$ and its negation, $-2 - 7i$.

Visually, these two vectors are shown as:

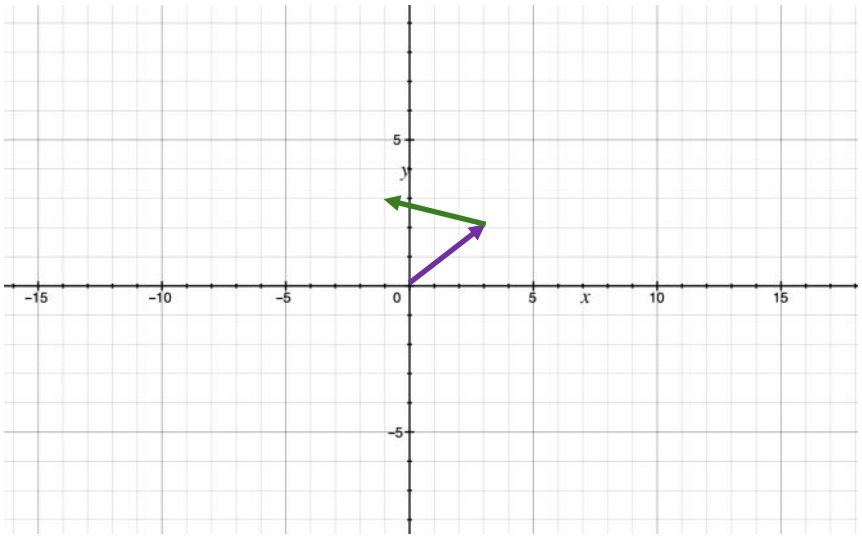


8.2.4 Example with Geometric Interpretation

Let us consider the complex number subtraction $(3+2i)-(4-i)$.

To subtract $(4-i)$ from $(3+2i)$, we find the vector difference of the two points using the tip to tail method. We recall a difference $(A-B)$ can be rewritten as a sum, $(A+(-B))$, which indicates a difference of vectors is found by adding the minuend vector to the negation of the subtrahend vector.

That is, $(3+2i)-(4-i)=(3+2i)+(-4+i)$



Observe the terminal point of the tip to tail combination lies at $(-1, 3)$, corresponding to the complex number $-1 + 3i$.

8.2.5 The Powers of *i*

Prerequisite for understanding the multiplication and division of complex numbers, we first examine the powers of the imaginary unit. As we will see, the powers of i follow a cyclical pattern, a fact we can exploit.

Recall any non-zero value raised to the power zero is equal to 1. Also, any value raised to the power one is simply the value. Therefore, $i^0 = 1$, while $i^1 = i$.

If we take a value to the second power, we are multiplying that value by itself. In this instance, $i^2 = -1$, by definition, and so $i^3 = (i^2)(i^1) = (-1)i = -i$.

Continuing, we encounter a surprising result: $i^4 = (i^2)(i^2) = (-1)(-1) = 1$. The fourth power of i has returned to the value found by raising i to the power zero, and therefore, the succession of values $1, i, -1, -i$, will continue for all powers of i. Ultimately, we will refer to this behavior as a cyclic pattern of order 4; that is, the values cycle through and then repeat values in a progression of four distinct values.

We can use the Remainder Theorem and the Power and Product Rules of Exponents from algebra to reveal a useful fact about i^K:

If $\frac{K}{4} = Q + \frac{R}{4}$, where Q is the quotient and R the remainder, then $K = 4Q + R$:

$$i^K = i^{4Q+R} = i^{4Q}e^R = (i^Q)^4 e^R$$

We have seen $i^4 = 1$, and so

$$i^K = i^R$$

Example 8.10 Find i^{77}.

Solution: Division shows us $\frac{77}{4} = 77 \div 4 = 19 + \frac{1}{4}$, and therefore, $i^{77} = i^1 = i$.

8.3 Multiplication and Division of Complex Numbers

To multiply two complex numbers, we can use the distributive property or the binomial multiplication technique FOIL (first, outer, inner, last) and the fact that $i^2 = -1$. These two operations reinforce our previous statement that, although complex numbers are not linear expressions, we can computationally treat them as though they were, with the proviso we remember the result must be completely simplified. Consider two complex numbers $z_1 = a + bi$ and $z_2 = c + di$.

The product of z_1 and z_2 is:

$$\begin{aligned} z_1 \times z_2 &= (a + bi)(c + di) \\ &= ac + adi + bci + bdi^2 \\ &= ac + adi + bci + bd(-1) \\ &= ac + adi + bci - bd \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

Example 8.11 Multiply the complex numbers $(2+3i)$ and $(4-i)$.

Solution: To multiply two complex numbers, we apply the FOIL process and follow the rules for algebraically combining imaginary numbers.

$$\begin{aligned}(2+3i)\ ^*\ (4-i) &= (2\ ^*\ 4) + (2\ ^*\ -i) + (3i\ ^*\ 4) + (3i\ ^*\ -i)\\ &= 8 - 2i + 12i - 3i^2\\ &= 8 + 10i - 3(-1)\\ &= 8 + 10i + 3\\ &= 11 + 10i\end{aligned}$$

Therefore, the product of $(2+3i)$ and $(4-i)$ is $11+10i$.

Example 8.12 Multiply the complex numbers $(3+2i)$ and $(4-i)$.

To multiply two complex numbers, we use the FOIL process and follow the rules for imaginary numbers.

$$\begin{aligned}(3+2i)\ ^*\ (4-i) &= (3\ ^*\ 4) + (3\ ^*\ -i) + (2i\ ^*\ 4) + (2i\ ^*\ -i)\\ &= 12 - 3i + 8i - 2i^2\\ &= 12 + 5i - 2(-1)\\ &= 12 + 5i + 2\\ &= 14 + 5i\end{aligned}$$

Therefore, the product of $(3+2i)$ and $(4-i)$ is $14+5i$.

There is a geometric interpretation of the product of complex numbers, but it is not as evident as the interpretation of addition and subtraction.

When considering a complex number $a+bi$, there are two significant components which must be taken into consideration: the magnitude of the complex number, defined as $\sqrt{a^2+b^2}$, and the angle θ between the vector representing the complex number and the positive x-axis. The product of two complex numbers will have a magnitude equal to the product of the individual complex number magnitudes and an angle equal to the sum of the two angles described above.

We mention, at this time, only for completeness and will return to the notion shortly.

8.3.1 Division of Complex Numbers

Dividing complex numbers is only slightly more involved than multiplication, but it does involve the use of **conjugate values**. When an expression consists of two terms, A + B or A − B, the conjugate to the expression is the identical expression with the operation between the expressions reversed. We say the terms are joined in a reciprocal relationship; in this case, the term reciprocal refers to reversed, and the relationships involved are addition and subtraction.

Therefore, the conjugate of $x+2$ is $x-2$; the conjugate of $t-\sqrt{5}$ is $t+\sqrt{5}$.

We have encountered conjugates in the past, during the process of rationalization of rational expressions involving radicals in the denominator; in fact, the process here is precisely the same, suppressing the conversion of i to $\sqrt{-1}$.

Recall, as an illustration, the rationalization of a rational expression involving a square root in the denominator.

Example 8.13

$$\frac{4+\sqrt{3}}{5-\sqrt{7}}$$

To rationalize means to remove the square root from the denominator; this is done by multiplying the numerator and denominator by the conjugate of the denominator and simplifying completely.

The conjugate of $5-\sqrt{7}$ is $5+\sqrt{7}$.

$$\frac{4+\sqrt{3}}{5-\sqrt{7}} \cdot \frac{5+\sqrt{7}}{5+\sqrt{7}}$$

In both the numerator and the denominator, we simplify; you will observe the square roots in the denominator subtracted from the problem.

$$\frac{20+4\sqrt{7}+5\sqrt{3}+\sqrt{21}}{25+5\sqrt{7}-5\sqrt{7}-\sqrt{49}}$$

$$\frac{20+4\sqrt{7}+5\sqrt{3}+\sqrt{21}}{25-7}$$

$$\frac{20+4\sqrt{7}+5\sqrt{3}+\sqrt{21}}{18}$$

$$\frac{10}{9}+\frac{2}{9}\sqrt{7}+\frac{5}{18}\sqrt{3}+\frac{1}{18}\sqrt{21}$$

None of the remaining square roots simplifies, and therefore, this is the rationalized form of the original expression.

When we say the square roots do not simplify, we mean the radicands have no perfect square factors; a square root such as $\sqrt{40}$ simplifies as $\sqrt{4}\cdot\sqrt{10}=2\sqrt{10}$.

To divide one complex number by another, we can apply this same notion of the conjugate of a complex number, now specifically referred to as a **complex conjugate**.

The complex conjugate of a complex number $z=a+bi$ is denoted as $z^*=a-bi$. Likewise, the complex conjugate of $a-bi$ is $a+bi$. Recall that $i=\sqrt{-1}$ and consider the following:

Example 8.14 Divide $\frac{6-4i}{5+3i}$.

Solution: Regarding the expression as $\frac{6-4\sqrt{-1}}{5+3\sqrt{-1}}$, we see the conjugate of the denominator is $5-3\sqrt{-1}$, which is the same as $5-3i$. Therefore, the numerator and the denominator will be multiplied by this expression. Note the conversion back and forth between i and $\sqrt{-1}$ is unnecessary, and so we will suppress that annotation in this and future calculations.

$$\frac{6-4i}{5+3i}\cdot\frac{5-3i}{5-3i}$$

$$\frac{30-18i-20i+12i^2}{25-15i+15i-9i^2}$$

Using the definition $i^2=-1$, we continue:

$$\frac{30-38i-12}{25+9}$$

$$\frac{18-38i}{34}$$

$$\frac{9}{17}-\frac{19}{17}i$$

Example 8.15 Divide the complex number $(5 + 3i)$ by $(2 - i)$.

Solution:

$$(5 + 3i)/(2 - i)$$

$$\frac{5 + 3i}{2 - i} \cdot \frac{2 + i}{2 + i}$$

$$\frac{10 + 5i + 6i + 3i^2}{4 + 2i - 2i - i^2}$$

$$\frac{10 + 11i - 3}{4 + 1}$$

$$\frac{7 + 11i}{5}$$

$$\frac{7}{5} + \frac{11}{5}i$$

8.3.2 Important Algebraic Properties

The algebraic properties of real numbers apply (almost) directly to the complex numbers. While we use these relationships in a manner almost subconscious, it is wise to present them formally here: Commutative property of addition and multiplication $z_1 + z_2 = z_2 + z_1$ and $z_1 \times z_2 = z_2 \times z_1$.

Associative property of addition and multiplication: $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$ $(z_1 \times z_2) \times z_3 = z_1 \times (z_2 \times z_3)$.

Distributive property: $z_1 \times (z_2 + z_3) = (z_1 \times z_2) + (z_1 \times z_3)$, and $z_1 \times (z_2 - z_3) = (z_1 \times z_2) - (z_1 \times z_3)$.

Observe nothing is mentioned in the commutative and associative properties about division and subtraction. As was the case for real numbers, these operations are neither commutative nor associative.

8.4 Graphical Representation of a Complex Number

As mentioned, the complex plane is a two-dimensional coordinate system, where the horizontal axis represents the real part of a complex number, and the vertical axis represents the imaginary part. Any complex number $z = a + bi$ can be plotted as a point (a, b) on this plane. The horizontal axis is called the real axis, the vertical axis the imaginary axis.

8.4.1 Magnitude and Argument

Each complex number $z = a + bi$ can be represented by two important quantities: the magnitude (or modulus) and the argument (or angle).

Magnitude (Modulus): The magnitude of a complex number $z = a + bi$, denoted as $|z|$, is the distance from the origin $(0, 0)$ to the point (a, b) on the complex plane. It is calculated as:

$$|z| = \sqrt{(a\text{^}2 + b\text{^}2)}$$

Observe this is simply the distance formula, describing the distance from the point (0, 0) to the point (a, b).

Argument (Angle): The argument of a complex number $z = a + bi$, denoted as arg(z), is the angle between the positive real axis and the line segment from the origin (0, 0) to the point (a, b) on the complex plane.

When calculating the argument of a complex number, there is a choice to be made between taking values in the range $(-\pi, \pi]$ or the range $[0, 2\pi)$. Both are equivalent and equally valid. We will use the convention $-\pi < \theta \leq \pi$.

For arguments in the span $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, we may calculate the angle θ as:

$$\theta = \arg(z) = \tan^{-1}(b/a)$$

This formula is valid only for angles in the span mentioned above, and so care should be taken for complex numbers whose coordinate representation lies in the second or third quadrants. The argument for those numbers having representation in those quadrants should be found using $\left(\tan^{-1}\frac{b}{a}\right) - \pi$.

The nuance being applied here may appear picayune, but is essential. Consider the complex numbers $3 + 4i$ and $-3 - 4i$. Each is easily seen to have a modulus equal to 5; what are their arguments?

Using the formula $\arg(z) = \tan^{-1}(b/a)$, we find (using a calculator) 0.927 (approximately). Since it is easier to visualize angles in degrees (until we are familiar with radians more precisely), the angle produced by the calculator is approximately 53.13°.

For the complex number 3 + 4i, this is correct. However, the complex number $-3 - 4i$ has argument $0.927 - 3.14 = -2.213$.

The reason for this is the peculiar and precise definition of the inverse tangent function, which deserves reinforcement here. When we say $\theta = \tan^{-1}\frac{b}{a}$, we mean θ is the unique angle in the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ such that $\tan\theta = \frac{b}{a}$. That interval of values describes the first and second quadrants only! Greater caution is needed when the angle is meant to be in either the second or the third quadrant, when an angle of $\pi/2$.

8.4.2 Polar Form of a Complex Number

A complex number $z = a + bi$ can also be expressed in polar form, which uses the magnitude and argument:

$$z = |z| (\cos(z) + i \sin(z))$$

This representation is particularly useful for performing operations such as multiplication and division of complex numbers, as well as for understanding the behavior of complex numbers in various applications.

Example 8.16 Find the magnitude and argument of the complex number $z = -2 + 3i$.

Solution: The magnitude (modulus) of $z = -2 + 3i$ is:

$$|z| = \sqrt{((-2)\text{^}2 + 3\text{^}2)} = \sqrt{(4 + 9)} = \sqrt{13}$$

The argument (angle) of $z = -2 + 3i$ is:

$$\arg(z) = \tan\text{^}-1(3/-2) = \tan^{-1}(-3/2) \approx -0.9828 \text{ radians (or approximately } -56.31°).$$

Example 8.17 Find the magnitude and argument of the complex number $(3 + 4i)$ in polar form:

To convert a complex number to polar form, we use the formulas:

$$\text{Magnitude (r)} = \sqrt{(a\text{^}2 + b\text{^}2)}$$
$$\text{Argument } (\theta) = \tan^{-1}(b/a)$$

where a is the real part, and b is the imaginary part of the complex number.

In this case, we have:

$$a = 3$$
$$b = 4$$

Calculating the magnitude:

$$\begin{aligned} \text{r} &= \sqrt{(3\text{^}2 + 4\text{^}2)} \\ &= \sqrt{(9 + 16)} \\ &= \sqrt{25} \\ &= 5 \end{aligned}$$

Calculating the argument:

$$\begin{aligned} \theta &= \tan^{-1}(4/3) \\ &\approx 0.9273 \text{ radians} \\ &\approx 53.13^\circ \end{aligned}$$

Therefore, the complex number $(3 + 4i)$ in polar form is:

$$\text{Magnitude (r)} = 5$$
$$\text{Argument } (\theta) = 53.13^\circ.$$

Example 8.18 Represent $(-2 - 3i)$ in polar form.

To convert a complex number to polar form, we use the formulas:

$$\text{Magnitude (r)} = \sqrt{(a\text{^}2 + b\text{^}2)}$$
$$\text{Argument } (\theta) = \tan^{-1}(b/a)$$

where a is the real part, and b is the imaginary part of the complex number.In this case, we have:

$$a = -2$$
$$b = -3$$

Calculating the magnitude:

$$\begin{aligned} \text{r} &= \sqrt{((-2)\text{^}2 + (-3)\text{^}2)} \\ &= \sqrt{(4 + 9)} \\ &= \sqrt{13} \\ &\approx 3.6056 \end{aligned}$$

Calculating the argument:

$$\begin{aligned} \theta &= \tan^{-1}(-3/-2) \\ &= \tan^{-1}(3/2) \\ &\approx -0.9828 \text{ radians} \\ &\approx -56.31^\circ \end{aligned}$$

Since the argument is negative, we can add 360° to get the final answer in the range of 0 to 360°.

Therefore, the complex number $(-2-3i)$ in polar form is:

Magnitude (r) ≈ 3.6056
Argument $(\theta) \approx 303.69°$.

Example 8.19 Represent $(0+5i)$ in polar form.

To convert a complex number to polar form, we use the formulas:

Magnitude (r) = $\sqrt{(a\text{^}2 + b\text{^}2)}$
Argument $(\theta) = \tan^{-1}(b/a)$

where a is the real part, and b is the imaginary part of the complex number.

In this case, we have:

$a = 0$
$b = 5$

Calculating the magnitude:

$$\begin{aligned} \text{r} &= \sqrt{(0\text{^}2 + 5\text{^}2)} \\ &= \sqrt{25} \\ &= 5 \end{aligned}$$

Calculating the argument:

$$\begin{aligned} \theta &= \tan^{-1}(5/0) \\ &= \pi/2 \text{ radians} \\ &= 90° \end{aligned}$$

Therefore, the complex number $(0+5i)$ in polar form is:

Magnitude (r) = 5
Argument $(\theta) = 90°$.

Consider two complex numbers, $1+i$ and $-1-\sqrt{3}$i, and let us find their product. Using the method of multiplication, we have seen previously, we obtain $(-1+\sqrt{3}) - (\sqrt{3}+1)i$. Now, we establish facts about each of the two complex numbers and their product.

For $1+i$, the modulus of the complex number is $\sqrt{2}$, and the argument is $\frac{\pi}{4}$, while $-1-\sqrt{3}$i has modulus 2 and argument $-\frac{2\pi}{3}$.

The product has modulus $2\sqrt{2}$ and argument $-\frac{5\pi}{12}$.

We now notice an interesting relationship: the modulus of the product is the product of the moduli of the two complex numbers, and the argument of the product is the sum of the arguments of the two complex numbers.

8.5 Exponential Form

Closely related to the polar form of a complex number is the exponential form. The exponential form of a complex number $z = a + bi$ is given by:

$$z = |z|\, e^{i\theta}$$

where $|z|$ is the magnitude (modulus), and θ is the argument (angle) of the complex number.

The exponential form leverages the properties of exponential and trigonometric functions and the considerable computational machinery developed for them. Specifically, we can rewrite the polar form using Euler's identity:

$$z = |z|(\mathbf{cos}(\theta) + i\mathbf{sin}(\theta)) = |z|\ \mathbf{e}^{i\theta}$$

The identity can be derived in several ways, including an application of power series and techniques of integral calculus. Since we have examined series already, let's establish Euler's identity using that approach.

The series expansion for $e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$. Recalling the powers of i are cyclical (i.e., $i^0 = 1$, $i^1 = i$, $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, …) and taking the power on e to be $z = ix$, where x is an arbitrary complex number, we find:

$$e^{ix} = \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \cdots$$

Using the powers of i mentioned a moment ago, this becomes:

$$e^{ix} = \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} = 1 + ix - \frac{(x)^2}{2!} - \frac{i(x)^3}{3!} + \frac{(x)^4}{4!} + \cdots$$

The power series expansion of e^z is known to be absolutely convergent, and therefore, we can create a rearrangement of the series terms without affecting the sum. In particular, let us use the rearrangement where the real and imaginary parts are grouped:

$$e^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots\right)$$

Note the real part of this expression is precisely the power series representation of cosine, while the imaginary part is the power series representation of sine.

The conclusion is precisely the equality represented in the Euler identity.

Example 8.20 Express the complex number $z = 2 + 3i$ in exponential form.

Solution: The exponential form of a complex number is given by:

$$z = \mathrm{r} * \mathrm{e}^{i\theta}$$

where

r is the magnitude (or modulus) of the complex number,

θ is the argument (or angle) of the complex number, measured in radians,

To convert the given complex number $z = 2 + 3i$ to exponential form, we need to first find the magnitude (r) and argument (θ).

Step 1: Calculate the magnitude (r)

$$\mathrm{r} = \sqrt{(a\text{^}2 + b\text{^}2)}$$
$$\mathrm{r} = \sqrt{(2\text{^}2 + 3\text{^}2)}$$
$$\mathrm{r} = \sqrt{(4 + 9)}$$
$$\mathrm{r} = \sqrt{13}$$
$$\mathrm{r} \approx 3.6056$$

Step 2: Calculate the argument (θ)

$$\theta = \tan^{-1}(b/a)$$

$\theta = \tan^{-1}(3/2)$

$\theta \approx 0.9828$ radians

$\theta \approx 56.31^\circ$

Now, we can express the complex number $z = 2 + 3i$ in exponential form:

$z = \mathrm{r} * \mathrm{e}^{i\theta}$

$z = 3.6056 * \mathrm{e}^{i0.9828}$

$z = 3.6056 * (\cos(0.9828) + i * \sin(0.9828))$

Example 8.21 Convert the given complex number $z = -1 + \sqrt{3}i$ to polar form.

Solution:
Step 1: Calculate the magnitude (r)

$\mathrm{r} = \sqrt{(a\text{^}2 + b\text{^}2)}$

$\mathrm{r} = \sqrt{((-1)\text{^}2 + (\sqrt{3})\text{^}2)}$

$\mathrm{r} = \sqrt{(1 + 3)}$

$\mathrm{r} = \sqrt{4}$

$\mathrm{r} = 2$

Step 2: Calculate the argument (θ)

$\theta = \tan^{-1}(b/a)$

$\theta = \tan^{-1}(\sqrt{3}/-1)$

$\theta = \tan^{-1}(-\sqrt{3})$

$\theta \approx -1.0472$ radians

$\theta \approx -60^\circ$

Now, we can express the complex number $z = -1 + \sqrt{3}i$ in polar form:

$z = \mathrm{r} * (\cos(\theta) + i * \sin(\theta))$

$z = 2 * (\cos(-1.0472) + i * \sin(-1.0472))$

8.5.1 Applications of Complex Numbers

Complex numbers have many important applications in various fields of mathematics and science, including:

- Electrical engineering: Complex numbers are used to represent and analyze alternating current (AC) circuits, which involve both real and imaginary components.
- Quantum mechanics: Complex numbers are essential in the mathematical formulation of quantum mechanics, where they are used to represent the wave function and probability amplitudes.
- Geometry and transformations: Complex numbers can be used to represent and perform geometric transformations, such as rotations and reflections, in the complex plane.
- Signal processing: Complex numbers are used in the analysis and processing of signals, particularly in the context of Fourier analysis and the study of complex-valued functions.
- Fractals and chaos theory: Complex numbers are fundamental to the study of fractals and chaotic systems, where they are used to generate and analyze intricate geometric patterns.

Applications of Complex Numbers as Vectors

The vector representation of complex numbers has numerous applications, including:

- Representing and transforming 2D coordinates and images.
- Analyzing electrical circuits and signals.
- Describing and analyzing oscillations and waves.
- Describing and analyzing rotation and other geometric transformations.

8.6 Applications of Complex Numbers in Cybersecurity

Complex numbers have become an increasingly important tool in the field of cybersecurity, with applications ranging from cryptography to network analysis.

One of the primary applications of complex numbers in cybersecurity is cryptography. Many modern encryption algorithms, such as the Rivest–Shamir–Adleman (RSA) cryptosystem, rely on the properties of complex numbers to ensure the security of sensitive data.

In RSA, the public and private keys are generated using the factorization of large prime numbers. These prime numbers can be represented as complex numbers, and their properties, such as the ability to generate unique factors, are leveraged to create secure encryption and decryption methods.

Additionally, complex numbers are used in the implementation of elliptic curve cryptography (ECC), which is an alternative to traditional RSA-based encryption. Elliptic curves can be defined over the complex number plane, and the mathematical operations performed on these curves provide a robust foundation for secure communication.

Complex numbers also play a crucial role in the analysis and optimization of communication networks, which are the backbone of modern cybersecurity infrastructure.

In network analysis, the impedance of a circuit, which is the measure of the opposition to the flow of alternating current (AC), can be represented using complex numbers. This representation allows for the accurate modeling of network behavior, including the identification of potential bottlenecks, signal attenuation, and phase shifts.

9

Exponentials and Algorithms

9.1 Exponential Functions

Exponential functions exhibit a rate of change—the change per instant or unit of time—proportional to the function's current value. If that rate of change is positive, the situation is described as exponential growth, while a negative change is described as exponential decay.

These functions are characterized by a constant base that is raised to a variable power. When we say a "variable power," we refer to a situation where the exponent is an expression involving a variable, such as x. We shall find these functions play a crucial role in modeling a wide range of phenomena in various fields, such as finance, biology, chemistry, and physics.

The general form of an exponential function is:

$$f(x) = \mathbf{a} \cdot \mathbf{b}^{\boldsymbol{x}}$$

where

- a, the constant coefficient of the exponential, also happens to be the y-intercept of the graph, or the initial value of the function.
- b is the base of the exponential, a non-negative real number not equal to 1, whose value determines the rate of growth or decay.
- x is the variable exponent in the general form.

9.1.1 Properties of Exponential Functions When Stated in General Form

1) Increasing or decreasing: Exponential functions either increase or decrease throughout their domain, depending on the value of the exponential base b:
   - If $b > 1$, the function is increasing (exponential growth).
   - If $0 < b < 1$, the function is decreasing (exponential decay).
2) Asymptotic behavior: Exponential functions of the form $f(x) = a * b^x$ have a horizontal asymptote on the left at $y = 0$ for increasing functions.
3) Rapid growth or decay: Exponential functions exhibit rapid growth or decay, especially as the independent variable x increases.
4) Inverse relationship with logarithmic functions: Exponential functions and logarithmic functions are inverse functions of each other. This relationship is crucial for solving equations involving exponential and logarithmic expressions.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

9.1.2 The Effect of the Value of the Base

Whether exponential functions in general form represent growth or decay is determined entirely by the value of the non-negative base of the exponential. If that base is greater than 1, we have exponential growth; if that base is between 0 and 1, we have exponential decay.

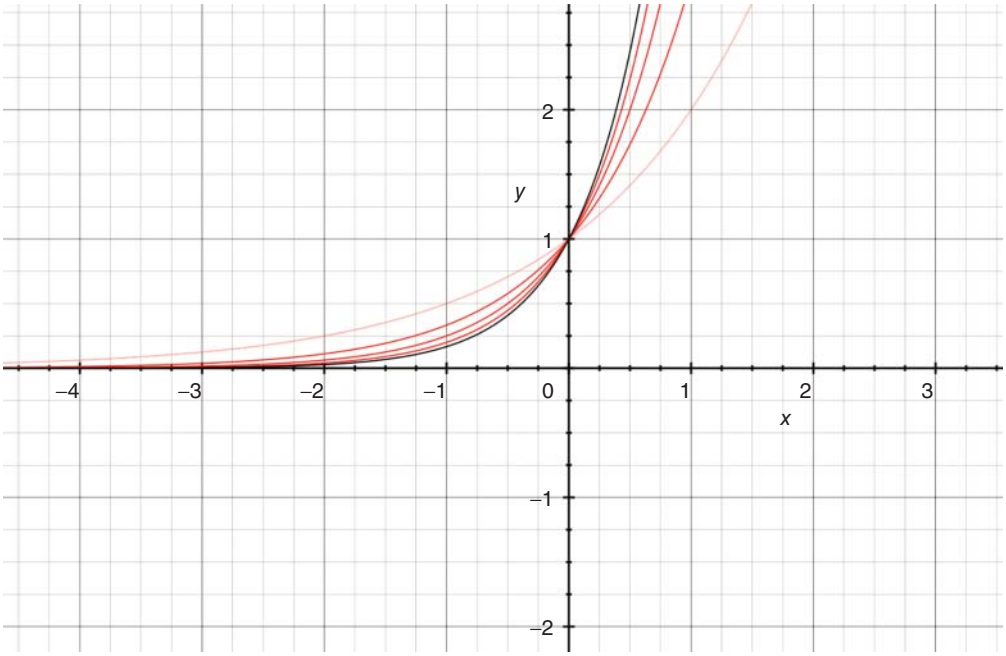
To see why this is the case, we consider a table of values used to determine values of exponential functions with various bases, beginning with the case of exponential growth.

We will consider the specific situation where $f(x) = b^x$.

base\exponent	−4	−3	−2	−1	0	1	2	3	4
2	0.062	0.125	0.25	0.5	1	2	4	8	16
3	0.012	0.037	0.111	0.333	1	3	9	27	81
4	0.003	0.015	0.062	0.25	1	4	16	64	256
5	0.001	0.008	0.04	0.2	1	5	25	125	625
6	0.000	0.004	0.027	0.166	1	6	36	216	1296

Observe characteristic features of the table. First, when the exponent is 0, all the exponential values shown are equal to 1; this is due to the zero-exponent rule. Second, the function values approach 0 from above as the values of x become progressively more negative; this is due to the negative exponent rule. Third, the function values rise rapidly as we increase the values of x.

We provide the graphs of these examples of exponential growth below:

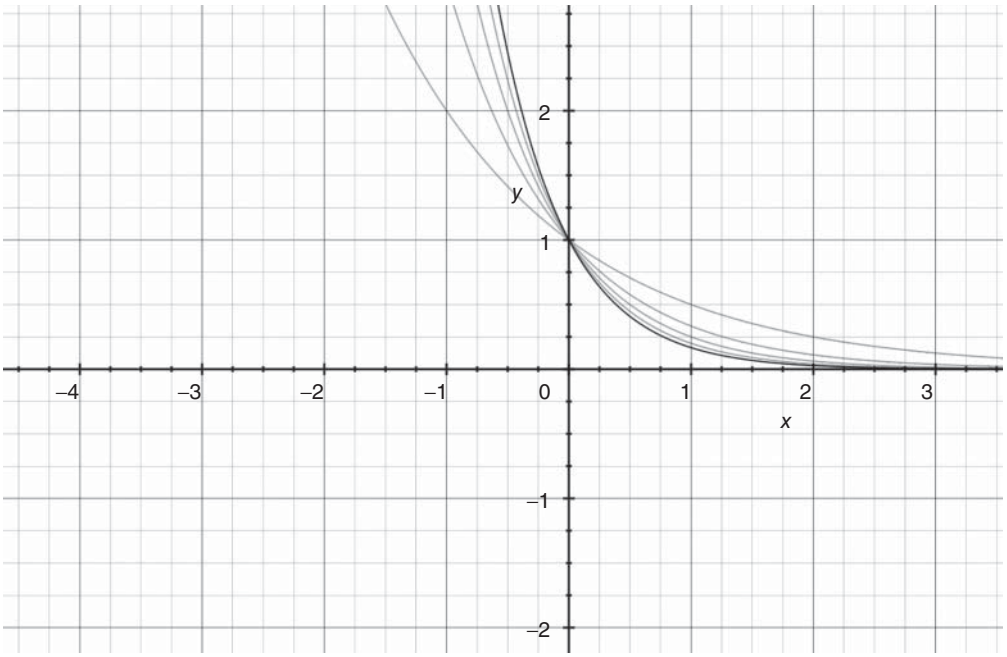


The greater the value of the base b, the more rapid the progression of the curve toward the x-axis on the left, and the more rapid the increase of the curve to the right.

Next, consider exponential decay graphs.

base\exponent	−4	−3	−2	−1	0	1	2	3	4
1/2	16	8	4	2	1	0.50	0.25	0.13	0.06
1/3	81	27	9	3	1	0.33	0.11	0.04	0.01
1/4	256	64	16	4	1	0.25	0.06	0.02	0.00
1/5	625	125	25	5	1	0.20	0.04	0.01	0.00
1/6	1296	216	36	6	1	0.17	0.03	0.00	0.00

The values of the powers of 1/4, 1/5, and 1/6 in the lower right area of the table are shown rounded to the nearest hundredth. They are not, in fact, zero, but very small decimal numbers. We show the graphs of these functions below:



The smaller the value of the base b, the more rapid the increase of the curve on the left, and the more rapid the decrease of the curve toward the x-axis to the right.

9.1.3 The Common Base of Exponentials

As has been mentioned, the base of an exponential function can be any positive real number except for 1. There are two bases deserving of particular attention: 10, which we call the common base, and e, which we call the natural base. The value of e will be discussed in detail momentarily.

Because our number system is base-10, and because of its ease of exponentiation, we refer to 10 as the common base. We know, for instance, that 10 raised to some power (positive or negative) is found as a decimal number involving a "1" either preceded by a given number of zeros (for negative powers of 10) or followed by a given number of zeros (for positive powers of 10).

Let us first remind ourselves of the positive powers of 10. When 10 is raised to a positive integer power k, the result is a 1 followed by k zeros. That is, $10^5 = (10)(10)(10)(10)(10) = 100{,}000$. Since we know this to be the case, we do not need to write the intermediate step showing the individual values of 10 being multiplied together.

Negative powers of 10 are a bit more subtle, illustrated best through an example in which we include a detailed explanation.

Example 9.1 Write the decimal form of 10^{-3}.

Solution:
The definition of negative exponents tells us $10^{-3} = \frac{1}{10^3} = \frac{1}{1000} = 0.001$.

The format of the solution may contain a superfluous zero in the ones place, just to the left of the decimal point. Students frequently write such values in a manner that omits that zero, and it is quite normal to do so. However, in many fields of study, this is viewed as entirely unacceptable, as it could lead to mistakes whose consequences are devastating and life-threatening.

In the medical field, for instance, the value 0.1 milliliters, or 0.1 mL (when referencing an amount of medication to be delivered to a patient), written without the zero preceding the decimal point, would appear as .1 mL. It is, as one can imagine, possible the decimal point could be overlooked, and the patient given 1 mL of medication. Observe the patient has now been given a dose of medicine 10 times greater than was intended, and this could easily prove fatal.

For clarity, it is vital the zero preceding the decimal point be shown. It turns out this suits our purpose well in this instance, as it permits us to expand the rule for positive powers of 10 in such a way we have a convenient method for expressing all the powers of 10, positive or negative.

The Powers of 10 in Decimal Form:

- If 10 is raised to a positive integer power K, the result is a 1 followed by K zeros.
- If 10 is raised to a negative integer power K, the result is a 1 preceded by K 0's followed by a 1, with the decimal point placed to the right of the first of those zeros.

Example 9.2 Write the following powers of 10 in decimal form.

a) 10^4
b) 10^{-6}
c) 10^{-2}
d) 10^7

Solution:

a) Since the power on 10 is positive 4, we know the result is a one followed by four zeros, or 10,000.
b) Since the power on 10 is negative 6, we know the result is six zeros followed by a one, with a decimal point following the first zero. That is, 0.000001.
c) Since the power on 10 is negative 2, we know the result is two zeros followed by a one, with a decimal point following the first zero. That is, 0.01.
d) Since the power on 10 is positive 7, we know the result is a one followed by 7 zeros, or 10,000,000.

9.1.4 The Natural Base of Exponentials

There are numerical values referred to as "irrational," meaning they have no fraction representation involving integers. Such values have endless decimal representations that follow no discernible pattern or repetition.

An illustration of such a number is π, the ratio of the circumference of any circle to its diameter. Regardless of the size of the circle, the computational value of this ratio is always the same and is designated by the Greek letter π.

Another illustration of such a number is $\boldsymbol{e}$, which we refer to as the natural base of exponentials. The term "natural" is applied to this base because it arises naturally in the mathematical study of virtually every application (engineering, finance, optics, acoustics, geology, and so forth).

The natural base is defined to be the limit as x tends to infinity of the expression $\left(1+\frac{1}{x}\right)^x$. Preliminary examination of this expression reveals that (as x becomes infinitely large) the expression tends toward 1^∞, which is problematic (in Calculus, this form is called "indeterminate"). We know 1 raised to any power is coincident with 1, but the base of this exponential never actually achieves the value of 1 (in the limit, it becomes arbitrarily close to 1, but never becomes 1).

A table of values will reveal what happens to this expression for relatively small values of x:

x	$\left(1+\frac{1}{x}\right)^x$
1	2
2	2.25
3	2.37037037
4	2.44140625
5	2.48832
6	2.52162637
7	2.5464997
8	2.56578451
9	2.58117479
10	2.59374246
20	2.65329771
40	2.68506384
80	2.70148494
160	2.70983558
320	2.71404664
640	2.71616121
1280	2.71722076
2560	2.7177511
5120	2.71801642
10 240	2.71814911
20 480	2.71821547

Observe that, as x increases in size, we see it is tending toward (at least!) 2.718; if we were to continue increasing the values of x (remember, we have a long way to go if we are allowing the

value of x to "race off" to infinity!), we will begin to see the numbers in the decimal expansion matching from row to row; the limiting value to which those values approach is e.

An effective approximation for e is 2.71828, but the full decimal expansion is endless, never forming a repeating pattern of digits.

9.1.5 Applications of Exponential Functions

As has been mentioned, numerous applications of exponential functions exist. We present several here and introduce more in the exercises.

Simple Interest: (A non-exponential illustration for the purpose of review)
There are two types of interest earned by monetary investments: simple and compound. The term "simple" is not synonymous (in this case) with "easy" but instead refers to interest that is not compound.

Simple interest obeys the rule $I = PRT$, where:

- I refers to the interest earned
- P refers to the initial investment, or Principal
- R refers to the decimal form of the interest rate
- T refers to the time in years

The accumulated value A of the account is found by adding the principal to the interest. That is,

- $A = P + I$

Example 9.3 An investment of \$9500 is made in an account earning 3.8% simple interest for 18 months. What is the amount of interest earned, and what is the accumulated value of the investment?

Solution:
Using the simple interest formula, we see $P = \$9500$, $R = 0.038$, $T = 1.5$. Remember, the value of R is the decimal form of the interest rate, and T is the time in years. We need to make appropriate conversion of units in order to use the simple interest formula correctly.

The amount of interest earned is $(\$9500)(0.038)(1.5) = \$541.50$, and the accumulated value of the investment is $\$9500 + \$541.50 = \$10{,}041.40$.

Compound Interest
When we refer to compound interest, we mean an investment that actually earns simple interest, with a slight modification to the process. The interest earned is periodically added to the existing value of the investment, changing the value of the principal P in the simple interest formula when that addition occurs. The new total then earns interest for that same period of time, and the process repeats itself. There are numerous types of compound interest, including:

- Daily: 365 compounding periods in one year
- Monthly: 12 compounding periods in one year
- Quarterly: 4 compounding periods in one year
- Semi-Annually: 2 compounding periods in one year
- Annually: 1 compounding period in one year
- Continuously: An unlimited number of compounding periods in one year

The number of compounding periods in one year is referred to as "n."

The formula for the accumulated value (also called the future value) of an investment earning a given percentage compound interest (for all types other than continuously compounded interest) is an exponential function having the form:

$$A = P(1 + r/n)^{nt}$$

- A is the future (or accumulated) value of the investment
- P is the principal investment
- r is the interest rate, in decimal form
- n is the number of times the interest is compounded per year
- t is the time in years

Example 9.4 Suppose \$9500 is invested for 18 months at 3.8% interest, compounded daily. What is the accumulated value of the investment?

Solution:
This example incorporates variable values of an example we considered for simple interest, so some of the values may be familiar. We see $P = \$9500$, $r = 0.038$, $n = 365$, and $t = 1.5$. Therefore,

$$A = (9500)\left(1 + \frac{0.038}{365}\right)^{(365)(1.5)}$$
$$A \approx (9500)(1.000104109)^{547.5}$$
$$A \approx 10{,}057.20$$

The accumulated value of the investment is \$10,057.20.

Observe the difference in the accumulated value (as compared to the simple interest scenario) is only \$15.80; this might make it seem the distinction between simple and compound interest is a trivial matter. The power of compound interest becomes evident only after significant time has elapsed (in the example, only 18 months had transpired).

> *Note: It is understood that in all problems pertaining to currency, the solution values are rounded to two decimal places, reflecting dollars and cents. The dollar sign precedes the numerical value, despite the apparent contradiction with the spoken form (where the word "dollars" does not come first!).*

Example 9.5 Suppose you invest \$9500 in an account earning 3.8% interest compounded daily for 45 years. What is the accumulated value of the investment? What would the accumulated value of the investment have been had simple interest been earned?

Solution:
Before the computations begin, let us remind ourselves of what is happening. The investment amount is deposited, and at the end of the first day (since there is daily compounding), interest is calculated and added to the principal. During the second day, the process repeats itself, with a new interest amount added to the balance. Every day, the amount of interest will be greater than that of the preceding day, because the principal is growing.

Using the compound interest formula, with $P = \$9500$, $r = 0.038$, $n = 365$, $t = 45$, we find:

$$A = (\$9500)\left(1 + \frac{0.038}{365}\right)^{16425}$$
$$A = \$52{,}520.45$$

Had the interest earned been simple interest, we would deduce:

$$I = PRT$$
$$I = (\$9500)(0.038)(45)$$
$$I = \$16{,}245$$

Therefore, the accumulated value is:

$$A = P + I$$
$$A = \$9500 + \$16{,}245$$
$$\$25{,}745$$

Over the course of 45 years, the difference is a staggering \$26,775.45 in accumulated interest! It is the quality of time passing that allows the power of compound interest to be seen.

Imagine the same problem is approached using an even faster rate of compounding: hourly. A trivial computation reveals there are 8766 hours in one year.

Example 9.6 Suppose you invest \$9500 in an account earning 3.8% interest compounded hourly for 45 years. What is the accumulated value of the investment? How is this different from the result we obtained when interest was compounded daily?

Solution:
Before the computations begin, let us remind ourselves of what is happening. The investment amount is deposited, and at the end of the first hour (since there is hourly compounding), interest is calculated and added to the principal. Each hour, the process repeats itself, with a new interest amount added to the balance. Every hour, the amount of interest will be greater than that of the preceding hour, because the principal is growing.

Using the compound interest formula, with $P = \$9500$, $r = 0.038$, $n = 8766$, $t = 45$, we find:

$$A = (\$9500)\left(1 + \frac{0.038}{8766}\right)^{394470}$$
$$A = \$52{,}524.93$$

Note a curious result; even though we have increased the frequency of compounding from daily to hourly (there are now 24 times more opportunities for interest to compound), the difference after 45 years is \$4.48. What's interesting is there seems to be a "diminishing return" as the frequency of compounding increases. This leads us to the notion of taking that frequency to its ultimate limit, where interest is compounded, not daily nor hourly nor minute-by-minute, nor second-by-second, but at every single instant.

Continuously Compounded Interest
Interest that is "continuously compounded" is (theoretically) calculated every instant, added to your balance, and the process repeats. This is the fastest possible growth for an investment and is an illustration of exponential growth.

Despite its impressive description, over relatively short periods of time, the accumulation will not be much greater than daily compounding.

The formula for continuously compounded interest is an exponential function having the form:

$$A = Pe^{rt}$$

- A is the future (or accumulated) value of the investment
- P is the principal investment

- e is the natural base
- r is the interest rate, in decimal form
- t is the time in years

Why must the process of increasing the frequency of compounding lead to this new expression that involves e?

The formula for compound interest is:

$$A = P\left(1 + \frac{r}{n}\right)^{nt}$$

If we perform some mathematical simplification on this, we obtain some interesting results. Recalling $\lim_{u\to\infty}\left(1 + \frac{1}{u}\right)^{u} = e$, we can make the substitution $u = \frac{n}{r}$. Observe that this indicates that $ur = n$, so the compound interest formula changes to

$$A = P\left(1 + \frac{1}{u}\right)^{urt}$$

or,

$$A = P\left(\left[1 + \frac{1}{u}\right]^{u}\right)^{rt}$$

Since $u = \frac{n}{r}$, as $n \to \infty$, we have $u \to \infty$, and so $P\left(\left[1 + \frac{1}{u}\right]^{u}\right)^{rt}$ approaches $A = Pe^{rt}$.

Example 9.7 Suppose you invest $9500 in an account earning 3.8% interest compounded continuously for 45 years. What is the accumulated value of the investment?

Solution:

$A = 9500$, $r = 0.038$, and $t = 45$. Therefore,

$$A = 9500e^{(0.038)(45)} \approx 52{,}525.13$$

After 45 years, the final balance is $52,525.13, which differs from the hourly compounding total only by about 20 cents. As you can see, there is a limit to how rapidly compound interest can accumulate.

Radioactive Decay

The decay of radioactive materials follows an exponential function:

$$N(t) = N_0 * e^{-\lambda t}$$

- $N(t)$ is the amount of radioactive material remaining at time t
- N_0 is the initial amount of radioactive material
- λ is the decay constant, which determines the rate of decay

One of the most well-known examples of radioactive decay is the decay of Carbon-14 (^{14}C), which is used in radiocarbon dating to determine the age of ancient organic materials.

Carbon-14 is a radioactive isotope of carbon. It is formed in the Earth's atmosphere when cosmic rays interact with nitrogen (^{14}N). This results in the conversion of nitrogen into carbon-14. Once formed, Carbon-14 is absorbed by living organisms in the same way as the more stable form of carbon, Carbon-12 (^{12}C). As long as the organism is alive, the ratio of ^{14}C to ^{12}C remains constant.

After an organism dies, it no longer absorbs Carbon-14, and the ^{14}C in its tissues starts to decay into Nitrogen-14 (^{14}N**)**. The decay process follows a half-life pattern, which means that half of the initial amount of Carbon-14 in a sample will decay into Nitrogen-14 every 5730 years.

This decay occurs through beta decay:

- A neutron in the Carbon-14 nucleus converts into a proton, emitting a beta particle (an electron) and an antineutrino in the process.
- This conversion transforms the Carbon-14 atom into a Nitrogen-14 atom.

Example 9.8 By measuring the remaining amount of Carbon-14 in a sample (for example, from a fossil or archaeological artifact), scientists can estimate how long it has been since the organism died. Since the decay rate is predictable, the age of the sample can be determined by comparing the current amount of ^{14}C to the initial amount when the organism was alive.

If a fragment of wood contains only 25% of the original Carbon-14, how old is the fragment of wood?

Solution:
We are unaware of how much Carbon-14 was present initially, but that does not matter for this question. We can choose a "convenient" amount, like 100 g. If we know 50% of the original amount is present after 5730 years, that will mean 50 g is present when $t = 5730$.

Using the half-life formula,

$$0.5\mathrm{N}_0 = \mathrm{N}_0{}^* (e)^{-5730\lambda}$$
$$0.5 = (e)^{-5730\lambda}$$
$$\ln 0.5 = (-5730\lambda) \ln e$$
$$\ln 0.5 = -5730\lambda$$
$$-\frac{\ln 0.5}{5730} = \lambda$$

This means the formula for the radioactive decay of Carbon 14 is $N(t) = N_0(e)^{\frac{\ln 0.5}{5730}t}$.

For our problem, at the time of interest, we have $N(t) = 0.25N_0$, so:

$$0.25N_0 = N_0(e)^{\frac{\ln 0.5}{5730}t}$$

Therefore,

$$0.25 = e^{\frac{t}{5730}}$$
$$\ln 0.25 = \frac{t}{5730}$$
$$2 = \frac{t}{5730}$$
$$11{,}460 = t$$

In 11,460 years, only 25% of the original amount of Carbon 14 is present, so the piece of wood is 11,460 years old.

Population growth: The growth of a population (over a relatively short span of time) can often be modeled using an exponential function:

$$\mathrm{P}(t) = \mathrm{P}_0{}^* e^{rt}$$

- $P(t)$ is the population size at time t
- P_0 is the initial population size
- r is the growth rate

Example 9.9 Over a short period of time, a population of a town grows exponentially according to the formula:

$$P = 18420e^{0.03t}$$

Solution:
What is the population at time $t = 5$ years? What is the population at $t = 10$ years? What is the population at $t = 20$ years?

Solution:
For $t = 5$, we find:

$$P = 18420e^{0.15} \approx 21{,}401$$

For $t = 10$, we find:

$$P = 18420e^{0.3} \approx 24{,}864$$

For $t = 20$, we find:

$$P = 18420e^{0.6} \approx 33{,}563.$$

Note the rate of increase of population is not constant. We can see this if we make a table:

year	population	increase
0	18 420	
5	21 401	2981
10	24 864	3463
20	33 563	8699

We see that the rate of increase is growing, which makes sense if we consider the way we know populations of towns change over time.

Over extended periods of time, populations do not demonstrate exponential growth but instead follow logistic growth. This is due to a phenomenon known as the "carrying capacity" of the environment, which precludes the limitless growth of a population in an exponential manner.

Exponential Equations
An exponential equation, as the name implies, is an equation that involves an exponential expression. In very rare cases, it is possible to solve such equations directly, using the following relationship:

If $a^x = a^y$, then $x = y$

Example 9.10 Solve the equation $3^{2x} = 81$.

Solution:
We observe $81 = 3^4$, so the equation is actually $3^{2x} = 3^4$, and so $2x = 4$, which tells us $x = 2$.

Example 9.11 Solve the equation $4^{3x-5} = 8$.

Solution:
Although 8 is not a power of 4, both 4 and 8 are powers of 2. Therefore, we can rewrite the equation as

$$2^{2(3x-5)} = 2^3$$
$$2(3x-5) = 3$$
$$6x - 10 = 3$$
$$6x = 13$$
$$x = \frac{13}{6}$$

The shortcoming of this method is evident: it is unlikely both sides of the equation will be powers of the same base. In a sense, this is like catching lightning in a bottle.

What, for example, would we do if confronted with the equation $5^{4x+3} = 15$? 15 is not a power of 5, and we cannot rewrite each side of the equation as powers of some other base. Using the knowledge currently at our disposal, we cannot solve the equation.

We could attempt to approximate a solution; we know $5^1 = 5$ and $5^2 = 25$, so the power of 5 must be somewhere between 1 and 2, possibly closer to 2, since 15 is closer to 25 than it is to 5. Perhaps, we could try

$$4x + 3 = 1.6$$

or,

$$4x = -1.4$$
$$x = -0.35$$

If this is the case, we would attempt to verify

$$4^{1.6} = 15$$
$$9.19 \approx 15$$

We have undershot our goal. Perhaps, we should try

$$4^{1.7} = 15$$
$$10.56 \approx 15$$

Again, we have undershot our goal. We try again:

$$4^{1.85} = 15$$
$$12.996 \approx 15$$

We are still short of our goal. Clearly, this approach, while beginning to bear fruit, is unacceptably tedious. We need a more precise, quicker way to find a solution to such an equation.

The answer to our dilemma lies in the introduction of a new idea, the logarithm.

9.2 Logarithms

The problem of solving exponential equations for which the rule $a^x = a^y$ implies $x = y$ does not apply, which has thus far defied our best efforts at solution, was first solved by John Napier (1550–1617), a Scottish mathematician. Recognizing these equations could not be solved using known techniques, he proposed a new concept that would allow us to find a solution, the **logarithm**.

As was the case with the definition of imaginary numbers, the definition of logarithms was a particular construct intended to solve precisely the sort of problem we are now facing. Its creation, however, led to a ripple effect of properties worthy of study and important in the solution of many other problems.

Definition of Logarithms
The logarithm of a number x with base b is the exponent "a" to which b must be raised to get x. That is, if $b^a = x$, then $a = \log_b x$. We read the right-hand side of this expression as "the logarithm base b of x," or more casually as "the log base b of x."

The equation $b^a = x$ is said to be in "exponential form," while $a = \log_b x$ is said to be in "logarithmic form." The two forms are mathematically equivalent to one another; we mean by this that they express one and the same notion.

Example 9.12 Express $3^4 = 81$ in logarithmic form.

Solution:
Comparing this equation to the definition of logarithms, we see $b = 3$, $a = 4$, and $81 = x$. Therefore, the equivalent logarithmic form is $4 = \log_3 81$.

Example 9.13 Express $7\left(\frac{2}{5}\right)^3 = \frac{8}{11}$ in logarithmic form.

Solution:
Observe the problem, as stated, does not allow us to immediately make the conversion. In the definition, the exponential is seen to be completely isolated, and therefore our first step must be to achieve this.

$$\left(\frac{2}{5}\right)^3 = \frac{8}{77}$$

Now, we can perform the conversion to logarithmic form,

$$3 = \log_{2/5}\left(\frac{8}{77}\right)$$

Example 9.14 Express $\log_2 512 = 9$ in exponential form.

Solution:
The conversion to exponential form relocates the logarithmic base to be a base for the exponent "9" on the right-hand side of the equation. That is, we obtain $512 = 2^9$.

Example 9.15 Express $\log_4 17 = 3t$ in exponential form.

Solution:
Here, we reverse the conversion process, obtaining $17 = 4^{3t}$.

Note some cleverness could be applied here (though it is not, at this point, needed) by observing the right-hand side of the equation can be written as $(4^3)^t = 64^t$. This would change the appearance of the exponential form to be $17 = 64^t$.

While presenting the solution in this form is a novel way of asking you to keep properties of exponentials in mind, it does not enhance the solution to our conversion problem in any practical manner, at least not at this point. The reason we say this is the expression that remains is to find the power to which 64 must be raised to obtain 17. This does not further our attempt to determine the value of the variable t, which will be our objective.

Clever consideration points us toward an approximate answer: we know $63^{1/3} = 4$, and therefore $64^{2/3}$ is 16. It seems reasonable that the power we are seeking is slightly greater than 2/3. As mentioned earlier, trial and error might get us close to the true value, but this is unacceptably slow and inconvenient.

An alternative will be offered shortly.

The most often used logarithms are the common logarithm (base 10) and the natural logarithm (base *e*), which are denoted as $\log(x)$ and $\ln(x)$, respectively. Scientific and graphing calculators have specific keys to produce decimal approximations for these two types of logarithms.

> Note: when writing natural logarithms by hand, it is useful to either use capital letters (LN(x)) or cursive writing ($ln(x)$) to distinguish the "L" from the letter "I."

The notations for these special types of logarithms are motivated by traditions adopted during the days of manual typeset. In mathematics textbooks, 1's and 0's are used extensively, and so it proved convenient to omit the base in the set of the common logarithm. Similar rationale applies to the natural logarithm (extensive use of the letter "e" in English writing), but the motivation is more reasonably derived from the Latin form of the term for natural logarithms, *logarithmus naturalis*.

Prior to the advent of computers, great effort was expended in the creation of logarithmic tables, which contained logarithm values to some number of decimal places. An example is the 1898 text *Logarithmic Tables,* by Professor George William Jones of Cornell University. Such works, made redundant with modern technology, are tomes, which represent the sum of tremendous intellectual effort and should not be dismissed in terms of importance.

Properties of Logarithms

Logarithms possess several important properties that make them useful in various applications and that make calculations involving logarithms significantly easier:

1) Product Rule: $\log_b(x * y) = \log_b(x) + \log_b(y)$
2) Quotient Rule: $\log_b(x / y) = \log_b(x) - \log_b(y)$
3) Power Rule: $\log_b(x^y) = y * \log_b(x)$
4) Special logarithmic values: $\log_b 1 = 0$, and $\log_b b = 1$
5) The logarithm of a power of the base: $\log_b(b^n) = n$
6) Change of Base Formula: $\log_b(x) = \log_a(x) / \log_a(b)$

In these formulas, it is assumed the values of b, x, and y are such that the logarithmic expressions are well-defined.

The left-hand side of the equations in these properties is referred to as "condensed form," while the right-hand side of the equations is referred to as "expanded form." It is particularly useful in differential calculus to exploit the use of expanded form in complex computations. Why this is the case is best left to a course devoted exclusively to the subject, and we will not pursue it here.

The Change of Base Formula is particularly useful for decimal approximation of logarithms of bases other than 10 or e, if we make a choice of one of those values for "a" in the formula. Examples will make this fact abundantly clear.

The properties listed allow us to simplify and manipulate logarithmic expressions. It is unclear at the outset just why this should be the case. Be assured, we will see this to be the case through numerous examples.

Example 9.16 Demonstrate the truth of the product rule by computing the value of $\log_{10}(24)$ using the product rule in two ways, then confirm the result using exponentiation.

Solution:
Use of a scientific calculator shows us an approximation to the value of this expression is (rounded to the nearest ten-thousandth) 1.3802. Keep in mind the meaning of this result; if true, then $10^{1.3802}$ should be approximately 24. A check with the calculator finds an approximation of $10^{1.3802}$ to be 23.9993, which (as we can see) is extremely accurate—the exponential result is only $\frac{7}{10{,}000}$ from 24.

Let us now examine the logarithmic expression using a factorization of 24. We know that $24 = 4 \cdot 6$, among other possible factorizations that exist.

Using the product rule, we have:

$$\log_{10}(24) = \log_{10}(6) + \log_{10}(4)$$
$$\log_{10}(6) \approx 0.7782$$
$$\log_{10}(4) = 0.6021$$

Therefore, $\log_{10}(24) = \log_{10}(6) + \log_{10}(4) = 0.7782 + 0.6021 = 1.3803$.

We notice the results are virtually the same, varying only in a minor issue of rounding to the nearest ten-thousandth.

Using exponentiation, we compare this result to the one found earlier:

$\log_{10} 24 = 1.3803$ (approximately) corresponds to $24 = 10^{1.3803}$ (approximately). Using a calculator, $10^{1.3803} = 24.0049$ (approximately), which is obviously very close to 24.

Example 9.17 Simplify $\log_5(x^2 * y^3)$ using the product rule and the power rule of logarithms.

Solution:
Using the product rule, we can rewrite the given expression as:

$$\log_5(x^{2*} y^3) = \log_5(x^2) + \log_5(y^3)$$

Now, we can use the power rule for logarithms:

$$\log_5(x^2) = 2 * \log_5(x)$$
$$\log_5(y^3) = 3 * \log_5(y)$$

Substituting these back into the original expression, we obtain:

$$\log_5(x^2 * y^3) = 2\log_5(x) + 3\log_5(y)$$

Note: Quite often we are asked to "expand a logarithm," but not told which properties of logarithms to use. The rule of thumb is that if you CAN use a property of logarithms, then you MUST use that property during the process of expansion.

Example 9.18 Evaluate $\log_3(27 * 81)$ without using a calculator.

Solution:
Using the product rule, we can rewrite this as:

$$\log_3(27 * 81) = \log_3(27) + \log_3(81)$$

Observe 27 and 81 are both powers of the base 3. That is, our expression simplifies to:

$$\log_3 3^3 + \log_3 3^4$$

Using property 5, this expression simplifies to

$$3 + 4$$

$$7$$

Example 9.19 Expand $\log_7(12 * x^2 * y^3)$ using the product rule and power rule; do not use decimal approximations.

Solution:
Using the product rule of logarithms, we can rewrite this as:

$$\log_7(12 * x^2 * y^3) = \log_7(12) + \log_7(x^2) + \log_7(y^3)$$

Using the power rule of logarithms:

$$\log_7(x^2) = 2 * \log_7(x)$$
$$\log_7(y^3) = 3 * \log_7(y)$$

Substituting these into the original expression, we find:

$$\log_7(12 * x^2 * y^3) = \log_7(12) + 2 * \log_7(x) + 3 * \log_7(y)$$

It is tempting to halt here and state that we have reached the solution—but have we? Remember, if a property of logarithms CAN be used, it MUST be used. You may feel we have reached the limit of our possible simplifications but observe $12 = 4 \cdot 3 = 2^2 \cdot 3$. This means the product and power rules can be used again.

$$\log 7(12 * x2 * y3) = \log_7 2^2 + \log_7 3 + 2\log_7 x + 3\log_7 y$$
$$= 2\log_7 2 + \log_7 3 + 2\log_7 x + 3\log_7 y$$

The point here is to remember your powers of observation are being tested at every stage of the game. You may wonder if you are meant to be able to observe 12 can be broken down in this manner? The answer to that question is "yes, you are." For relatively small numbers (less than, say, 100), we expect you can recognize a prime factorization of particular values.

Example 9.20 Expand $\log_6(13 * 9)$ using the rules of logarithms.

Solution:
We have $\log_6(13 * 9)$

Using the product rule, we can rewrite this as:

$$\log_6(13 * 9) = \log_6(13) + \log_6(9)$$

Observing $9 = 3^2$, we continue to simplify:

$$\log_6(13 \cdot 9) = \log_6(13) + \log_6(3^2)$$
$$\log_6(13 \cdot 9) = \log_6(13) + 2\log_6(3)$$

Example 9.21 Expand $\log_5(100/4)$ using the rules of logarithms. Use two different methods, one of which is simplification of the fraction prior to simplification.

Solution:

$\log_5(100/4) = \log_5(100) - \log_5(4)$ *Quotient Rule of Logarithms*

Next, realizing $100 = 10^2$ and $4 = 2^2$, we have

$\log_5(100) - \log_5(4) = \log_5(10^2) - \log_5(2^2) = 2\log_5 10 - 2\log_5 2$ *Power Rule of Logarithms*

Since $10 = (5)(2)$, we continue:

$= 2[\log_5((5)(2))] - 2\log_5 2$
$= 2[\log_5 5 + \log_5 2] - 2\log_5 2$ *Product Rule of Logarithms*
$= 2\log_5 5 + 2\log_5 2 - 2\log_5 2$ *Distributive Law*
$= 2\log_5 5$
$= 2(1)$ *Since $\log_b(b) = 1$*
$= 2$

Next, we approach the problem by first reducing the fraction within the logarithm.

$\log_5(100/4) = \log_5 25$
$= \log_5(5^2)$
$= 2\log_5 5$ *Power Rule of Logarithms*
$= 2(1)$ *Since $\log_b(b) = 1$*
$= 2$

Notice that (by being observant) we could have shortened the process of expanding the logarithm significantly by reducing the fraction within the logarithm at the outset. Also, knowledge and use of different rules could prove more efficient.

When we began as:

$\log_5(100/4) = \log_5 25$
$= \log_5(5^2)$
$= 2$ *Since $\log_b(b^n) = n$*

This highlights the importance of being attentive; the more you can observe, the shorter a problem can be. We will shorten our further examples by omitting the explanation of some of the steps we feel you should, by now, recognize.

Example 9.22 Expand $\log_8(x^2/y^3)$ using the properties of logarithms.

Solution:

$\log_8(x^2/y^3) = \log_8(x^2) - \log_8(y^3)$ *Quotient Rule of Logarithms*
$\log_8(x^2/y^3) = 2\log_8(x) - 3\log_8(y)$ *Power Rule of Logarithms*

Example 9.23 Expand $\log_7(18/6)$ without using the properties of logarithms.

Solution:

$\log_7(18/6) = \log_7(18) - \log_7(6)$ *Quotient Rule of Logarithms*
$= \log_7 6 + \log_7 3 - \log_7 6$ *Product Rule of Logarithms*
$= \log_7 3$

Example 9.24 Simplify $\log_3(x^3/[2y^2])$ using the quotient rule and power rule.

Solution:

$\log_3 (x^3/[2y^2]) = \log_3 (x^3) - \log_3 (2y^2)$ *Quotient Rule of Logarithms*
$\log_3 (x^3/(2y^2)) = 3 \log_3 (x) - (\log_3 (2) + 2 \log_3 (y))$ *Product Rule of Logarithms*
$\log_3 (x^3/(2y^2)) = 3 \log_3 (x) - \log_3 (2) - 2 \log_3 (y))$

Example 9.25 Calculate $\log_4(50/25)$ using the quotient rule.

Solution:
We have $\log_4(50/25)$
Using the quotient rule, we can rewrite this as:

$$\log_4(50/25) = \log_4 (50) - \log_4 (25)$$

Now, we can use the fact that $\log_4(50) = \frac{\log 50}{\log 4} \approx \frac{1.6897}{0.6021} \approx 2.8219$ and $\log_4(25) = \frac{\log 25}{\log 4} \approx \frac{1.39794}{0.6021} \approx 2.32193$.
Substituting these values, we get:

$$\log_4(50/25) \approx 2.82193 - 2.32193 = 0.50000$$

As a check, the original problem could have been rewritten by reducing the fraction $50/25 = 2$, obtaining:

$$\log_4 2 = \log_4 4^{1/2} = (1/2)\log_4 4 = \frac{1}{2}.$$

Example 9.26 Evaluate $\log_5(z^3)$ when $z = 2$.

Solution:
We have $\log_5(z^3)$, and we are given that $z = 2$.
Substituting $z = 2$ into the expression, we obtain:

$$\log_5(z^3) = \log_5 (2^3)$$

Now, using the power rule, we can rewrite this as:

$$\log_5 (2^3) = 3 \log_5(2)$$

Example 9.27 Evaluate $\log_6(x^5)$ when $x = 3$.

Solution:
We have $\log_6(x^5)$, and we are given that $x = 3$.
Substituting $x = 3$ into the expression $\log_6(x^5)$, we obtain:

$$\log_6(3^5)$$

Now, using the power rule, we can rewrite this as:

$$\log_6(3^5) = 5 * \log_6(3)$$

Example 9.28 Calculate $\log_7(49)$ using the change of base formula.

Solution:
We have $\log_7(49)$.

Using the change of base formula, we can rewrite this as:

$$\log_7(49) = \log_{10}(49)/\log_{10}(7)$$

We find $\log_{10}(49) \approx 1.6902$ and $\log_{10}(7) \approx 0.8451$.

Substituting these values, we get:

$$\log_7(49) = 1.6902/0.8451 = 2$$

Example 9.29 Evaluate $\log_8(16)$ using the change of base formula.

Solution:

We have $\log_8(16)$

Using the change of base formula, we can rewrite this as:

$$\log_8(16) = \log_{10}(16)/\log_{10}(8)$$

We find that $\log_{10}(16) \approx 1.2041$ and $\log_{10}(8) \approx 0.9031$.

Substituting these values, we obtain:

$$\log_8(16) \approx 1.2041/0.9031 \approx 1.3333$$

Example 9.30 Evaluate $\log_6(25)$ using the change of base formula.

Solution:

Using the change of base formula, we can rewrite this as:

$$\log_6(25) = \log_{10}(25)/\log_{10}(6)$$

We know that $\log_{10}(25) = 1.3979$ and $\log_{10}(6) \approx 0.7782$.

Substituting these values, we get:

$$\log_6(25) = 1.3979/0.7782 \approx 1.7963.$$

Applications of Logarithmic Functions

Richter Scale: Earthquakes (natural shuddering of the Earth's surface) are the sudden release of energy in the Earth's crust, leading to ground shaking. This energy release usually occurs along faults—cracks in the Earth's crust where blocks of rock have moved relative to each other. The Richter scale (established by Charles F. Richter, an American seismologist, in 1935) is used to measure the magnitude (severity) of earthquakes and is based on a logarithmic scale.

Although the Richter scale has been largely replaced by the **moment magnitude scale** (Mw) for more accurate measurements of large earthquakes, it remains a well-known and historically significant method for assessing earthquake magnitude.

The formula for the Richter scale is:

$$M = \log_{10}(A/(A_0\Delta)$$

- M is the magnitude of the earthquake
- A is the maximum amplitude of the earthquake's seismic waves recorded on a device called a seismograph, measured in micrometers (also known more frequently as microns)
- Δ is the distance between the seismograph and the epicenter of the earthquake, measured in degrees

- $A_0\,\Delta$ is a correction factor that adjusts for the distance between the seismograph and the epicenter, ensuring that the scale is consistent regardless of where the seismograph is located on the surface of the Earth

A simplified form of the equation describing the Richter scale is $M = \log_{10} \frac{A}{S}$, where S is the standard intensity of a (whose amplitude is $1\,\mu m = 10^{-4}$ cm). For the sake of simplicity, we will work with this form of the equation.

The significance of the logarithmic nature of the Richter scale is that an increase of 1 unit in M corresponds to a 10-fold increase in the amplitude of the seismic waves, and a release of more than 30 times the amount of energy by the earthquake.

Example 9.31 In 1906, the earthquake in San Francisco registered 7.9 on the Richter scale. In the same year, another earthquake was recorded in Ecuador that registered 8.8 on the Richter scale. How much stronger was the Ecuadoran earthquake than the one in San Francisco?

Solution:
For each of the situations, we have the magnitude of the earthquake, and the value of S (whatever that may be) is the same constant value. Therefore, our equations are:

$$7.9 = \log \frac{A_{sf}}{S}$$

and

$$8.8 = \log \frac{A_E}{S}$$

Here, A_{sf} is the intensity of the San Francisco earthquake, and A_E is the intensity of the Ecuadoran earthquake. Since the quake in Ecuador was more intense, we want the ratio $\frac{A_E}{A_{sf}}$.

In the first equation, we find:

$$(\log A_E - \log S) - (\log A_{sf} - \log S) = 8.8 - 7.9$$

$$\log A_E - \log A_{sf} = 0.9$$

$$\log \frac{A_E}{A_{sf}} = 0.9$$

$$\frac{A_E}{A_{sf}} = 10^{0.9}$$

$$\frac{A_E}{A_{sf}} \approx 7.94$$

The quake in Ecuador was nearly eight times as strong as the one in San Francisco.

Decibel scale: The decibel (dB) scale, which is used to measure the intensity of sound, is also a logarithmic scale. Recalling the prefix "deci" indicates one-tenth, we see the decibel is one-tenth of a Bel—the Bel is a unit that is seldom used but is the source of the name for this concept. The formula for the decibel scale is:

$$\text{dB} = 10 * \log_{10}(P/P_0)$$

P is the power of the sound wave.

P_0 is a reference power level, signifying the threshold level below which humans cannot hear.

Example 9.32 In 1969, the classic rock/blues band Led Zeppelin once played a concert at which their rendering of the song "Heartbreaker" registered at approximately 130 dB. How much stronger was this sound than that of the minimum threshold humans can hear?

Solution:
We are interested in the ratio $\frac{P}{P_0}$, and so, using the decibel formula,

$$130 = 10 \log \frac{P}{P_0}$$
$$13 = \log \frac{P}{P_0}$$
$$10^{13} = \frac{P}{P_0}$$

The version of the song was 10,000,000,000,000 times, or ten-trillion times as loud as the faintest noise a person can hear. This is as loud as a jackhammer at close proximity and can cause permanent hearing loss if exposure is prolonged.

pH scale: The pH scale, which is used to measure the acidity or basicity of a solution, is a logarithmic scale. The formula for the pH scale is:

$$\text{pH} = -\log_{10}[\text{H+}]$$

[H+] is the concentration of hydrogen ions in the solution.

As was the case for the Richter and decibel scales, the pH scale is logarithmic, essentially meaning the difference in 1 pH unit reflects a difference of 10 times in the concentration of hydrogen ions. The term pH originates from Latin and is an acronym for "*potentia hydrogenii*"—the power of hydrogen.

The more hydrogen ions there are, the more acidic the solution. These concentrations are really, really small fractions, like 1/10,000,000 for pure water. Because this denominator ranges drastically, it is convenient to use a logarithmic scale.

pH is usually between 0 and 14. Knowing the dependence of pH on [H+], our interpretation is as follows:

- If pH < 7, then the solution is acidic.
- If pH = 7, then the solution is neutral.
- If pH > 7, then the solution is basic.

Example 9.33 What is [H+] for an acid solution whose pH is 3.28?

Solution:
We need to solve the expression –log[H+] = 3.28 for [H+].

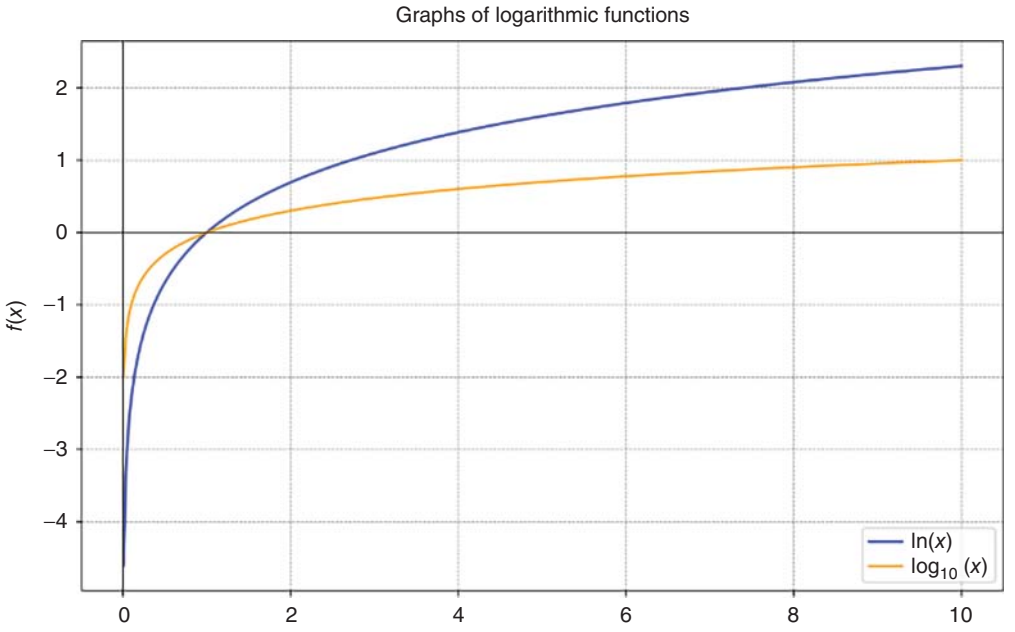
$$\log[\text{H+}] = -3.28$$

and so,

$$[\text{H+}] = 10^{-3.28}$$

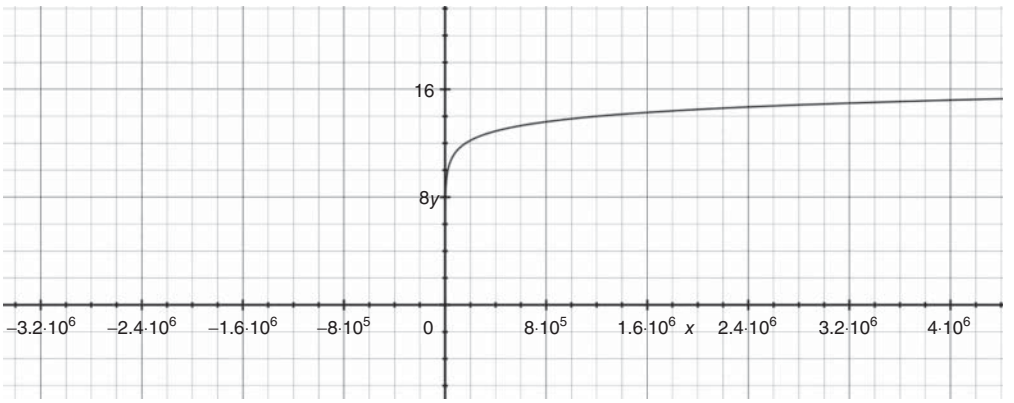
Using a scientific calculator, we find $[\text{H+}] \approx 0.000525$, or 5.25×10^{-4}.

Graphs of Logarithmic Functions
The graph of a logarithmic function $f(x) = \log_b x$ is a smooth, continuous curve that is increasing for $b > 1$ and decreasing for $0 < b < 1$. Logarithmic functions of the stated form have a vertical asymptote at $x = 0$.



In the graph above, the dark gray curve represents the natural logarithm function $\ln(x)$, and the orange curve represents the common logarithm function $\log(x)$. Notice that each of the functions has a vertical asymptote at $x = 0$. Although logarithmic functions appear to have a horizontal asymptote to the right, as well, this is deceptive.

Consider the graph of $f(x) = \ln(x)$ in a much wider window:



At first glance, we might suspect a horizontal asymptote on the right somewhere around $y = 16$. This is, however, untrue. Using calculus, it can be shown the graph of the natural logarithmic function (and all logarithmic functions) increases or decreases (depending on the value of the base b) incredibly slowly; yet they are always increasing or decreasing.

Thus, despite appearances, the graph of $f(x) = \ln(x)$ increases/decreases, ever so slowly, as the graph proceeds to the right.

9.3 The Properties of Exponential and Logarithmic Function Graphs

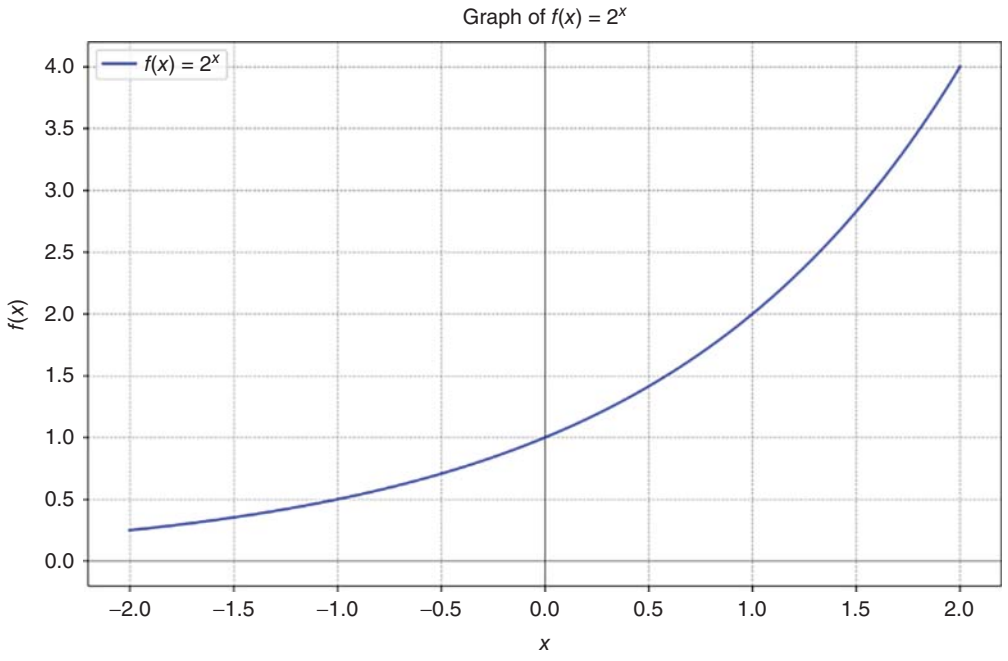
Extensive examination of the graphs of exponential and logarithmic functions lies beyond the scope of our work here. We will take a brief look at the graphs of the basic forms of these functions, however, and leave further consideration to the reader as a topic for review.

Exponential Function Graph Properties

1) Domain: The domain of an exponential function is all real numbers ($x \in \mathbb{R}$).
2) Range: The range of an exponential function with $a > 1$ is $(0, \infty)$. For $0 < a < 1$, the range is $(0, \infty)$.
3) Graphical behavior: The graph of an exponential function is a curve that either increases (if $a > 1$) or decreases (if $0 < a < 1$) as x increases.

Example 9.34 Graph the exponential function $f(x) = 2^x$, noting the domain and range.

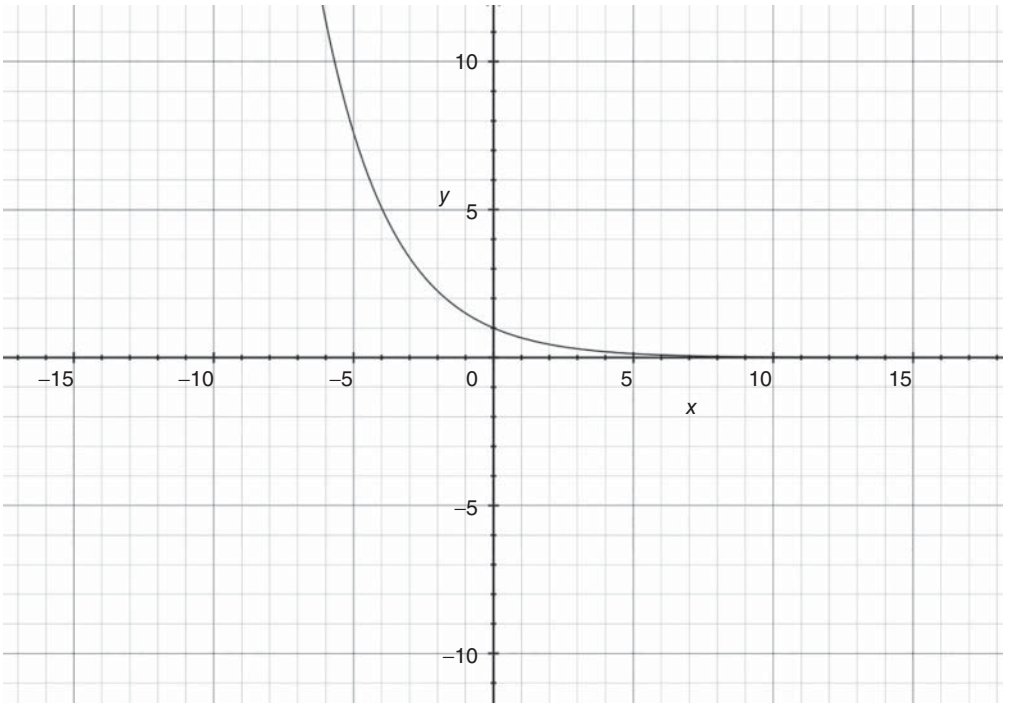
Solution:
Domain: $x \in \mathbb{R}$
Range: $(0, \infty)$
Graph:



Observe the graph of $f(x) = 2^x$ is a curve that increases, since the base is greater than 1.

Example 9.35 Graph the exponential function $f(x) = \left(\frac{2}{3}\right)^x$, noting the domain and range.

Solution:
Domain: $x \in \mathbb{R}$
Range: $(0, \infty)$
The graph of this function can be produced using a table of values or with a graphing utility:



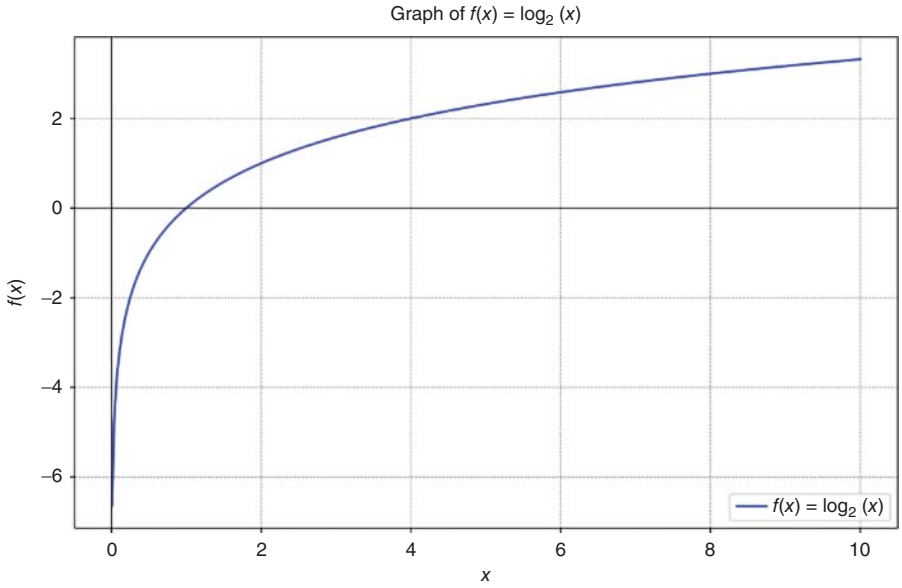
The graph is decreasing because the base is between 0 and 1.

Logarithmic Function Graph Properties

1) Domain: The domain of a logarithmic function with base a is $(0, \infty)$, since the logarithm of a non-positive number is undefined.
2) Range: The range of a logarithmic function is all real numbers ($\mathbb{R}$).
3) Graphical behavior: The graph of a logarithmic function is a curve that increases (if $a > 1$) or decreases (if $0 < a < 1$) as x increases.

Example 9.36 Find the domain and range of the logarithmic function $f(x) = \log_2(x)$ and produce its graph.

- Domain: $x \in (0, \infty)$.
- Range: $\mathbb{R}$.
- Graph:

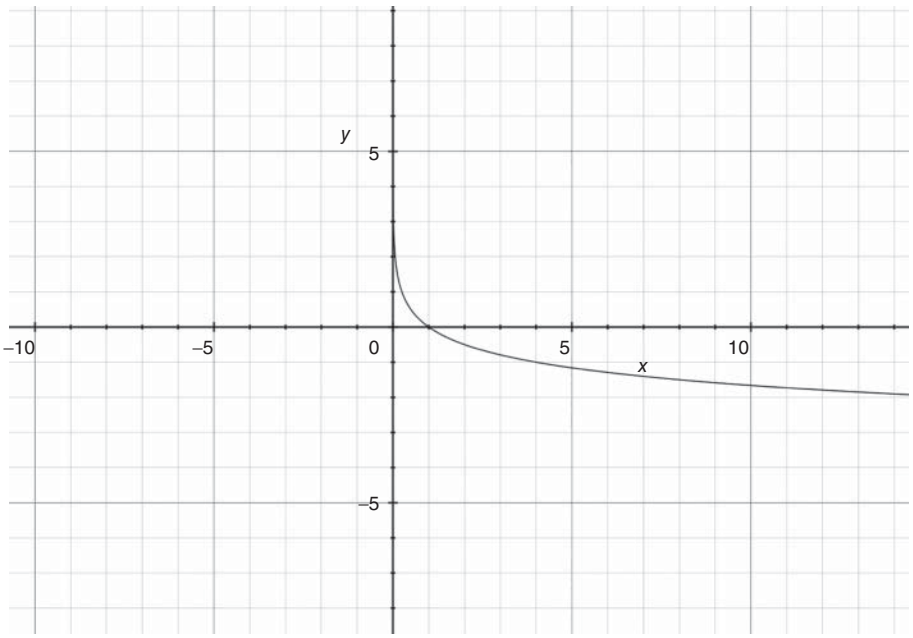


We see that, since the base is greater than 1, the function increases.

Example 9.37 Find the domain and range of the logarithmic function $f(x) = \log_{1/4} x$ and produce its graph.

Solution:

- Domain: $x \in (0, \infty)$.
- Range: $\mathbb{R}$.



Observe the graph decreases because the base is between 0 and 1.

9.4 Solving Exponential and Logarithmic Equations

9.4.1 Solving Exponential Equations

Exponential equations (that is, equations somehow involving exponential expressions) can be solved in one of two ways:

- Use of the identity stating that if $a^x = a^y$, then $x = y$
- Isolation of the exponential expression and subsequently converting to logarithmic form

We will first present an example solved in both ways.

Example 9.38 Solve the equation $3^{(2x+1)} = 27$ using both methods stated above and verify the solution.

Solution:
Observing the left-hand side of the equation presents a power of 3, we notice the right-hand side of the equation, 27, is also a power of 3. That is,

$$3^{2x+1} = 3^3$$

Equating the exponents, we find:

$$2x + 1 = 3$$
$$2x = 2$$
$$x = 1$$

Alternatively, to solve for x, we first need to isolate the exponential on the left-hand side of the equation. In this case, the exponential 3^{2x+1} is already isolated.

Now, we can use the definition that relates exponentials to logarithms:

$a^x = b$ if and only if $x = \log_a(b)$

Applying this property, we find:

$$2x + 1 = \log3(27)$$

or,

$$2x + 1 = \log_3 3^3$$
$$2x + 1 = 3$$
$$2x = 2$$
$$x = 1$$

Therefore, the solution to the equation $3^{(2x+1)} = 27$ is $x = 1$.

We next verify the solution:

Given equation: $3^{(2x+1)} = 27$
Substituting $x = 1$:
$3^{(2(1)+1)} = 3^3 = 27$

Example 9.39 Solve the equation $16^{(2x-3)} = 64$.

Solution:
Here, we can see 64 is not a power of 16, so it appears we must use the second alternative.

$$2x - 3 = \log_{16} 64$$

Interestingly, the Change of Base Theorem reveals the right-hand side of this equation to be equal to $\frac{\ln 64}{\ln 16} = 1.5$, exactly.
This means:

$$2x - 3 = 1.5$$

$$2x = 4.5$$

$$x = 2.25$$

At first glance, this is surprising. Our original equation did not allow us to use the property, stating that if $a^x = a^y$, then $x = y$, and this would lead us to believe the answer would be some expression involving logarithms.

However, under closer inspection, we realize both 16 and 64 are powers of 4. That is,

$$(4^2)^{2x-3} = 4^3$$

or,

$$4^{4x-6} = 4^3$$

Equating exponents,

$$4x - 6 = 3$$

$$4x = 9$$

$$x = \frac{9}{4}$$

$$x = 2.25$$

It is, admittedly, rare when the technique of equating exponents can be used. One might think, based on the first two examples, this is common. However, had the second example asked us to solve the equation $16^{(2x-3)} = 63$, our method would have failed; we would have been forced to convert to logarithmic form and solve there.

Example 9.40 Solve the equation $5^{(x+2)} = 600$.

Solution:
In this example, we notice 600 is not a power of 5, and therefore our only available method of attack is to convert to logarithmic form.

$$x + 2 = \log_5 600$$

$$x = -2 + \log_5 600$$

Checking the validity of the solution is a bit involved here, but we can satisfy ourselves of the reasonableness of the solution by using a decimal approximation, as $-2+\log_5 600 \approx -2+3.9746 \approx 1.9746$, and so by substitution, we find $5^{3.9746} \approx 599.965$, which is within thirty-five thousandths of 600; this gives us confidence our solution is valid.

9.4.2 Solving Logarithmic Equations

Before turning to the examples of solving logarithmic equations, we make an important observation: the expression $\log_b u$ is only defined if $u > 0$. While we can see this rather clearly in the examples above, that clarity lessens when we employ properties of logarithms while solving an equation.

This restriction on the permitted values in logarithmic expressions ***demands*** we check all our solutions to logarithmic equations for "validity," which means they are not in violation of this principle.

The primary technique used for solving logarithmic equations is the use of the definition,

$$\log_a u = v \text{ implies } u = a^v$$

It may be necessary to employ the properties of logarithms to apply this definition, but it is the general method for the solution.

If an equation were to involve exclusively logarithmic expressions, we can use the properties of logarithms to rewrite the equation as $\log_u v = \log_u w$, from which we can say $v = w$ and proceed from there.

Unless specifically instructed to use decimal approximation, the solution should be left in exact logarithmic form.

Example 9.41 Solve the equation $\log_4(x) = 2$.

Solution:
Here, we can immediately apply the definition of logarithms,

$$x = 4^2$$
$$x = 16$$

We verify the solution:

Substituting $x = 16$ into the original equation:

$$\log_4(16) = 2$$
$$\log_4 4^2 = 2$$
$$2\log_4 4 = 2$$
$$2 = 2$$

Therefore, the solution $x = 16$ satisfies the original equation.

Example 9.42 Solve the equation $\log_3(2x+1) = 2$.

Solution:
We can immediately apply the definition of logarithms, finding:

$$3^2 = 2x+1$$
$$9 = 2x+1$$

$$8 = 2x$$
$$4 = x$$

Let us verify the solution:
Substituting $x = 4$ into the original equation:

$$\log_3(2(4)+1) = \log_3(9)$$

$$\log_3 9 = \log_3 9$$

Example 9.43 Solve the equation $\log_3(2x+1) = \log_3(x+4)$.

Solution:
Here, we are given an equation in which one logarithm with base 3 is equal to another logarithm with base 3. Therefore,

$$2x+1 = x+4$$

So,

$$2x+1 = x+4$$
$$x+1 = 4$$
$$x = 3$$

We verify the solution by substitution into the original equation:

$$\log_3(2x+1) = \log_3(x+4)$$
$$\log_3(2(3)+1) = \log_3(3+4)$$
$$\log_3(7) = \log_3(7)$$

Therefore, the solution $x = 3$ satisfies the original equation and is considered valid.

Example 9.44 Solve the equation $\log_2(x) + \log_2(x-7) = 3$.

Solution:
To solve for x, we must use the product rule of logarithms and then the definition.

$$\log_2(x^2 - 7x) = 3$$

$$x^2 - 7x = 2^3$$

$$x^2 - 7x - 8 = 0$$

$$(x-8)(x+1) = 0$$

$$x = 8 \ \ or \ \ x = -1$$

We check each possible solution:

$$\log_2(8) + \log_2(8-7) = 3$$
$$3 + 0 = 3$$

Thus, $x = 8$ is valid.

$$\log_2(-1) + \log_2(-1-7) = 3$$

Each of the logarithms is undefined, so the result of $x = -1$ is invalid and discarded. Our solution to the equation is $x = 8$.

Example 9.45 Solve the equation $\log_5(x) + \log_5(2x - 1) = 3$.

Solution:
To apply the definition of logarithms, we must first combine the logarithms on the left-hand side of the equation using the properties of logarithms.

$$\log_5[x(2x - 1)] = 3$$

Then,

$$x(2x - 1) = 5^3$$

or,

$$2x^2 - x = 125$$

Recall polynomial equations are solved by first reorganizing so one side of the equation is zero, then employing factoring or other techniques to find a solution. Thus,

$$2x^2 - x - 125 = 0$$

Solving the quadratic equation here requires us to use the Quadratic Formula, as the trinomial does not factor.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(2)(-125)}}{2(2)}$$

$$x = \frac{1 \pm \sqrt{1 + 1000}}{4}$$

$$x = \frac{1 \pm \sqrt{1001}}{4}$$

The factorization of 1001 is (7)(11)(13), and so the square root does not simplify.

We observe the alternative $x = \frac{1-\sqrt{1001}}{4}$ to be negative, so we discard this as a possibility, as it would immediately make the first logarithm undefined.

The positive option gives us

$$\log_5\left(\frac{1 + \sqrt{1001}}{4}\right) + \log_5\left[2\left(\frac{1 + \sqrt{1001}}{4}\right) - 1\right] = 3$$

Each of the logarithms is defined, but it is unclear if the expression is "true." We will use decimal approximations and the change of base rule to determine if our result is numerically reasonable.

$$\frac{1 + \sqrt{1001}}{4} \approx 8.16$$

Thus, we verify

$$\log_5(8.16) + \log_5[2(8.16) - 1] \approx 3$$

$$1.304 + 1.696 \approx 3$$

$$3 \approx 3$$

There is some argument whether the use of a decimal approximation is appropriate as a check of a solution, but there are times we appeal to it to avoid unnecessarily tedious calculations that would take us far afield from the point of solving the equation. Having said this, the result of the check involving an approximation should be very, very close to precise equality to convince us of the validity of the solution.

> *Note: It is NOT simply the case that a negative solution is considered invalid BECAUSE it is negative*! *Logarithmic equations often have negative solutions. However, if a possible solution is checked in the original equation and causes one of the initial logarithms to be undefined, or if the resulting simplification proves numerically incorrect, THIS indicates that a solution being checked is invalid.*

9.5 Applications of Exponential and Logarithmic Functions in Cybersecurity

Exponential and logarithmic functions play a crucial role in various aspects of cybersecurity. These mathematical models provide powerful tools for understanding and addressing the dynamic nature of cyber threats, cryptographic algorithms, and network analysis. In this section, we will explore several key applications of exponential and logarithmic functions in the field of cybersecurity.

9.5.1 Cryptographic Algorithms

Cryptographic algorithms are the foundation of secure communication in the digital world. Many modern cryptographic algorithms, such as RSA and Diffie–Hellman, rely on the properties of exponential and logarithmic functions to ensure the confidentiality and integrity of data.

9.5.2 RSA Cryptography

The RSA cryptographic algorithm is based on the concept of public-key cryptography, which uses a pair of keys: a public key and a private key. The encryption and decryption processes involve exponential and logarithmic functions.

The public key is generated by raising a large prime number to a power, and the private key is derived from the public key using a logarithmic operation. The encryption process involves raising the plaintext to the power of the public key, while the decryption process involves raising the ciphertext to the power of the private key.

9.5.3 Network Security and Traffic Analysis

Exponential and logarithmic functions are also useful in the analysis of network traffic and the detection of cyber threats. The growth patterns of network traffic and the spread of malware often exhibit exponential characteristics, which can be analyzed to identify anomalies and potential security breaches.

9.5.4 Network Traffic Analysis

Consider the growth of a cyber-attack, such as a distributed denial-of-service (DDoS) attack, where the number of infected devices (bots) increases exponentially over time. The relationship between

the number of infected devices (N) and time (t) can be modeled using an exponential function:

$$N(t) = N_0 \cdot e^{rt}$$

where N_0 is the initial number of infected devices, and r is the growth rate of the infection. By analyzing the growth rate and the shape of the exponential curve, security analysts can detect and mitigate the DDoS attack in a timely manner (Figure 9.1).

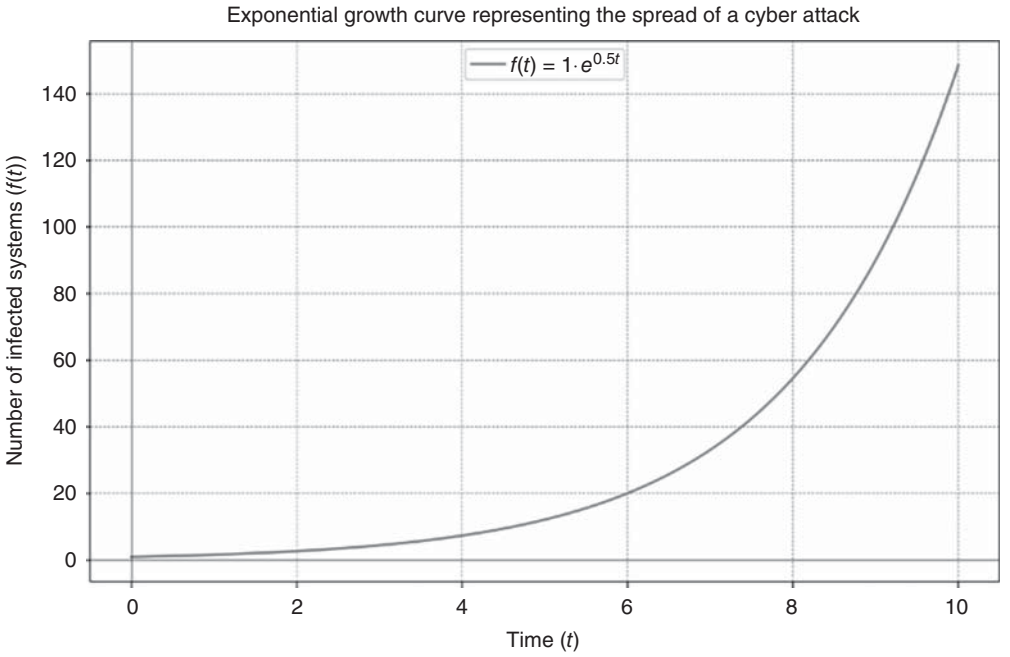


Figure 9.1 Shows an example of an exponential growth curve representing the spread of a cyber-attack.

9.5.5 Vulnerability Assessment and Risk Analysis

Logarithmic functions are often used in vulnerability assessment and risk analysis in cybersecurity. The severity of a vulnerability or the impact of a cyber incident can be measured on a logarithmic scale, which helps in prioritizing and addressing the most critical risks.

Example 9.46 CVSS (Common Vulnerability Scoring System)

The common vulnerability scoring system (CVSS) is a widely used framework for assessing the severity of software vulnerabilities. The CVSS score is calculated using a logarithmic formula that takes into account factors such as the exploitability of the vulnerability and the potential impact on the affected system.

The CVSS score ranges from 0 to 10, with higher scores indicating more severe vulnerabilities. Security professionals can use the CVSS scores to prioritize the remediation of vulnerabilities and allocate resources effectively.

9.5.6 Intrusion Detection and Anomaly Identification

Logarithmic functions can also be employed in intrusion detection systems (IDS) and the identification of anomalies in network traffic patterns. The logarithmic scale can be used to represent the rarity or significance of specific network events, which can help in the early detection of potential cyber threats.

Suppose that the typical network traffic in an organization follows a certain pattern, which can be modeled using a logarithmic function. By continuously monitoring the network traffic and comparing it to the expected logarithmic pattern, the IDS can detect any significant deviations or anomalies that may indicate the presence of a cyber-attack or unauthorized activity.

10

What Do We Mean by Statistics?

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect and summarize data.

Often, people think of probability and statistics interchangeably; this is actually untrue because they are two sides of a particular coin. Probability is a predictive science, using information to extrapolate toward future events. Statistics is a summative science, describing events that have already taken place.

The two concepts certainly do overlap, as statistical information can allow us to extrapolate toward future events, but they are not one and the same.

10.1 Sampling

Sampling is a crucial aspect of statistics that involves selecting (in some manner) a subset of individuals from a "population" to estimate characteristics of that population. There are several sampling techniques, each with its own advantages and disadvantages. In this section, we will discuss five common sampling methods: convenience sampling, systematic sampling, random sampling, cluster sampling, and stratified sampling.

An immediate question presents itself: why bother with sampling at all? If we are interested in information about a "population," why not just investigate the population?

The answer to this question is a matter of practicality. Typically, the population is so large and so far-flung that a direct examination of the population is impossible or impractical. It is for this reason that when the United States conducts a "Census" (a concept to which we shall return shortly), it is only done every ten years. Collection of data from the entire population is an extremely cumbersome and expensive proposition, and therefore, it is undertaken only periodically.

Countries with larger populations, such as China, have performed a census only sporadically. The People's Republic of China has performed a census only in the years 1953, 1964, 1982, 1990, 2000, 2010, and 2020.

The manner in which a sample is created dictates, in large part, the information gleaned from that sample. If poorly constructed, the sample will have limited utility as a tool for learning about the population; it could, in fact, lead to completely erroneous conclusions. We call such misinformation "sampling bias."

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

10.1.1 Convenience Sampling

Convenience sampling is a non-probabilistic sampling technique where samples are selected based on their ease (convenience) of availability. This method is often used when the population is large and resources are limited.

For example, a researcher might survey people at a local shopping mall to gather opinions about a new product. While convenience sampling is quick and inexpensive, it may not be representative of the entire population, leading to sampling bias. The reasons for this are manifold: first, participants must just happen to be patrons of the shopping mall on a given day and at a given time; second, the respondents are simply those who were willing to speak to the researcher; third, there is no reason to believe the respondents will be forthcoming with true answers, as they may wish to end their participation in the survey as quickly as possible. These and other reasons could lead to significant bias in the sampling result.

Example 10.1 Charismatic Senate candidate Maraca Kabob seeks input on the opinion of the citizens of his district prior to undertaking his campaign. Kabob unwittingly collects a sample of viewpoints at a store frequented by members of an opposing political party and finds (possibly to his surprise) the positions supported by the survey respondents are diametrically opposed to his own. Based on the results, Kabob rethinks his position on policies affecting the residents, causing support from within his own party to suffer. This is a convenience sample because Senatorial candidate Kabob chose to conduct his survey at a randomly chosen (and potentially hostile) location.

Example 10.2 A journalist wants to write an article about the favorite ice cream flavors among city residents. Due to time constraints, they decide to interview the first 50 people they encounter at a nearby park. This is an example of convenience sampling.

The rationale behind gathering a convenience sample lies at the heart of its potential flaws. There is no significant thought behind the structure of the sample, and therefore, we encounter an illustration of the popular adage, "garbage in, garbage out."

10.1.2 Systematic Sampling

Systematic sampling involves selecting elements from an ordered sampling frame at regular intervals. In effect, each element from the frame is assigned a positional reference number, and that order is immutable. The first element to be chosen for the sample is selected randomly, and then every kth element is selected from the list. The value of k is determined by dividing the population size by the desired sample size and rounding down to the nearest whole number; it may be necessary to "loop through" the positional reference numbers and return to the beginning in order to complete the sample.

Example 10.3 A company has a list of 1000 customers and wants to select a sample of 100 for a satisfaction survey. They choose a random starting point (perhaps the 205th customer in the list) and then select every 10th customer from the list ($k = 1000/100 = 10$). The members of the sample are the 205th, 215th, 225th, … 995th, 5th, 15th, …, 200th member.

Do you see how, when we reached and surpassed the population size, we were required to return to the beginning of the list in order to complete our selection of 100 members of the sample?

Example 10.4 A production line completes the assembly of 400 units of a particular product each hour. A systematic sample of 30 units is to be gathered for the purpose of inspection. Using a random

number generator, it is determined the first member of the sample will be the 47th unit produced in a given hour. Finding $400/30 = 13.33$ (approximately), we round down to 13 and begin selecting every 13th unit from the assembly line. The units selected are numbered 47, 60, 73, 86, …, 398, 11, 24, 37.

Why is this method of sampling "superior" to convenience sampling? Because the members of the sample are random (generated by the arbitrarily selected initial member number and sample size), the selection of sample members is unknown to the researcher at the outset.

Systematic sampling can be flawed, despite its superiority to convenience sampling.

Imagine a production process in which microprocessors are produced by a collection of 15 robotic assembly machines and the microprocessors are released sequentially. That is, the first machine produces a microprocessor, the second does likewise, and so on, in a rhythmic pattern. An arbitrary unit is chosen for initial inspection (suppose it is the 23rd unit off the line), and the algorithm for determining the frequency of units between choosing members of the sample produces a value of 30 (that is, every 30th unit of production is chosen for inspection).

Observe that, by coincidence, only units from the 8th robot in the production process are ever chosen to be a part of the sample. Though unexpected, the systematic sampling process has an inherent design flaw that precludes us from effectively analyzing the quality of the microprocessors created in this manufacturing process.

10.1.3 Simple Random Sampling

Simple random sampling is a probabilistic sampling technique where each member of the population has an equal chance of being selected. This method involves using a random number generator or a lottery system to select elements from the population. Random sampling minimizes bias and allows for the generalization of results to the entire population.

It is worth mentioning that a "random number generator" is actually a "mythical creature." Research has established that computer algorithms designed to produce random strings of digits are, in fact, incapable of producing truly "random" values. These are often referred to as "pseudo-random number generators," but it is common to still ascribe to them the label "random number generators."

The word "pseudo" is synonymous with the word "false" and indicates the resulting numbers (though they may appear to be random) are not truly random. Full exploration of this concept lies beyond the scope of our conversation here, but it is worth noting.

Example 10.5 A state wants to determine the average height of its school-aged students (5–18-years old). They assign each student a unique number and use a random number generator to select a series of 50 identification numbers. The members of the sample will correspond to those whose identification number matches one of the 50 generated values.

Under the assumption the numbers generated are truly random and not repeated, this methodology ensures each student has an equal probability of being chosen.

10.1.4 Cluster Sampling

Cluster sampling involves dividing the population into groups or clusters and then creating a simple random sample from the members of each cluster in order to create the sample. This technique is

useful when the population is geographically dispersed and when opinions of the individual regions may be at odds.

An illustration of the potential need for such a method of sampling can be seen in the challenges facing the state of Oregon in the current political climate. Significant numerical superiority of the Willamette Valley (predominantly liberal) creates a situation wherein the much more conservative and rural eastern portions of Oregon find themselves under-represented. This has led to the movement for secession of that eastern section of the state to form a proposed "Greater Idaho" region.

Example 10.6 A research team wants to study the level of infection of a rare disease in a particular country. They divide the country into regions, then create a simple random sample of the residents of these regions. The individuals selected within these regions form a cluster sample of the overall population of the country.

Cluster samples have the advantage of reducing the cost and time required in collecting information about the overall population. There is, additionally, the benefit of providing information about the variation that exists between the clusters.

The disadvantages of cluster sampling are that it does not produce a genuine "simple random sample," as well as the inherent challenge from all samples treated as a microcosm of the population: they are likely to be less representative of the population as a whole.

10.1.5 Stratified Sampling

"Strata" refers to the concept of "layers." In geology and related fields, a stratum is a layer of rock or sediment characterized by certain properties or attributes that distinguish it from adjacent layers from which it is separated.

In a similar manner, stratified sampling divides the population into subgroups (layers or strata) based on specific characteristics, such as age, gender, or income level. A random sample is then selected from each stratum in proportion to its size in the population. This method increases the likelihood that each subgroup is adequately represented in the sample.

Why would we want to achieve this representation?

Imagine a particular public policy proposal that was drafted in Congress, one which would impact individuals at various income levels in different and profound ways. An example of this might be the reformation of the Federal Income Tax system.

Because individuals at different levels of income would be impacted, it would be wise to determine the opinions of citizens at those levels. It would then create an atmosphere in which a balanced approach to such a policy might be undertaken.

Example 10.7 A political candidate wants to gauge public opinion on a new policy regarding conscription to the armed services. Opinions on this topic will, naturally, vary by age (younger individuals would likely be opposed to such a policy), and so the government opts to create a stratified sample to measure the pulse of the country on the issue. The population is stratified by age (18–34, 35–54, 55+), and individuals are randomly selected according to a proportional number model of individuals from each age group to participate in the survey.

Such a process ensures the voices of all individuals are given adequate consideration in the development of a policy on mandatory conscription. Like all sampling techniques, however, there is the

potential stratified sampling might prove problematic, particularly when individuals can fall into more than one of the strata.

An illustration of this is the notion of subdividing a population according to ethnicity. We live in an increasingly multicultural society, and so the classification of an individual into a particular ethnic category could prove problematic. One might suggest we allow individuals to "self-select" into a particular ethnic group, but this has the drawback of preventing a truly representative simple random sample be taken from each of the population strata.

10.2 Statistical Graphs

Statistical graphs are essential tools for visualizing and understanding data. They allow us to present complex information in a clear and concise manner, making it easier to identify trends, patterns, and relationships.

As is often the case, the tool is two-edged: the unscrupulous use of statistical graphs can mold public perception in such a manner that (though not specifically dishonest) does not precisely align with all the facts. This practice is sometimes referred to as the production of "misleading graphs." More will be said of this shortly, but once again, an adage is brought to mind: "Believe nothing of what you hear, and only half of what you see." (Edgar Allen Poe, paraphrase).

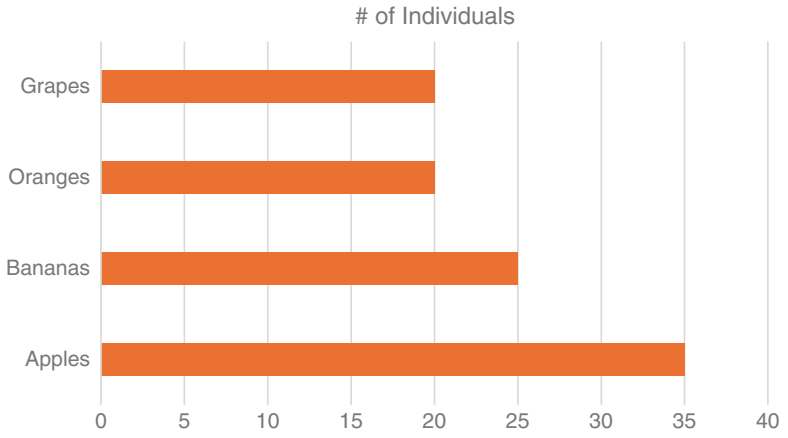
More will be said of misleading graphs after we have introduced each of the various types of statistical graphs.

10.2.1 Bar Graph

A bar graph is used to analyze categorical data. It consists of rectangular bars of equal width, with the height of each bar representing the frequency or value of the corresponding category. Bar graphs can be vertical or horizontal; if the bars are horizontal, the description just given must be modified: the "length" of each bar represents the frequency or value of the corresponding category.

Example 10.8 A survey was conducted to determine the favorite fruits among a group of 100 people. The results are as follows: 35 people prefer apples, 25 prefer bananas, 20 prefer oranges, and 20 prefer grapes (Figure 10.1).

Were the graph to be presented with horizontal bars (the choice of which type to use is generally a matter of esthetics or of which format is more persuasive), the graph would resemble:



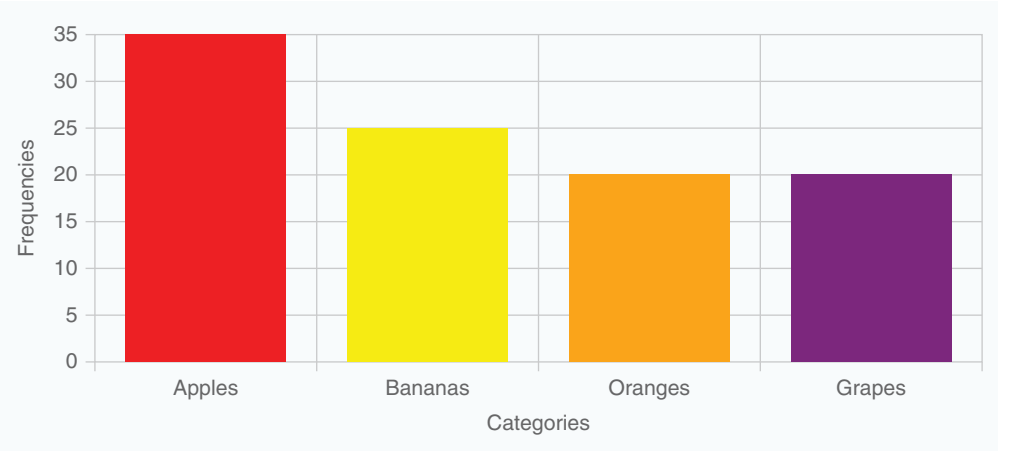
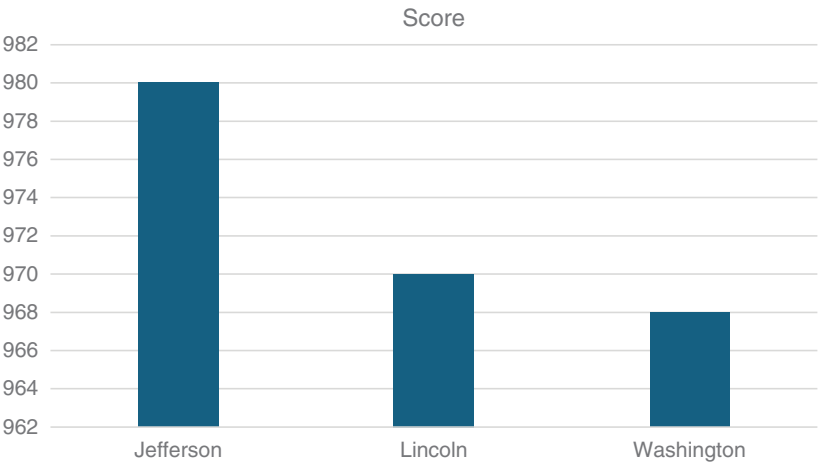


Figure 10.1 Favorite fruit bar graph.

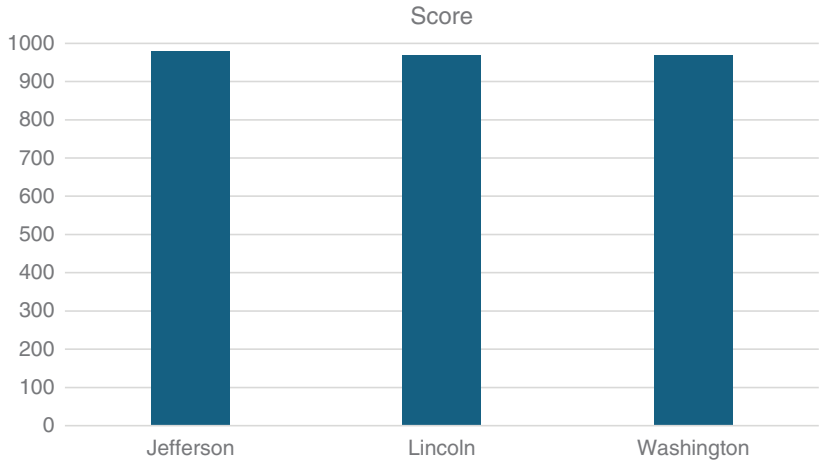
How can a bar graph be misused to present information designed to steer the casual observer toward a particular conclusion?

Consider the aptitude scores of senior-level students at three high schools in a particular district:



On casual inspection, the students at Jefferson School appear to be presenting significantly higher scores on the aptitude test. However, consider the scale of the vertical axis: it has been deliberately constructed in such a manner that we are led to a particular interpretation of the information.

Consider now another bar graph depicting the same information:



In this presentation, it appears the performance of students at the three schools is virtually identical.

The rationale for choosing one method of presentation over another is usually self-serving, in the sense that true information is intended to be presented but in the most favorable light. The first of the previous two graphs might be presented by the representatives of Jefferson School to emphasize the (albeit slightly) higher performance of their students over those of the other schools. Lincoln and Washington Schools would likely utilize the last graph shown as a means to emphasize the school students' performance among the three options is virtually the same.

Note, however, one important observation: neither of the presentations of data graphically is false! This is an important realization: an accusation of deliberate falsehood is a rather serious charge—it is incumbent upon the viewer of the information to understand the information being presented to them.

10.2.2 Pie Chart

A pie chart is a circular graph divided into sectors, where each sector represents a proportion of the whole. The proportion within the pie chart for each sector is determined by multiplying the percentage of each sector by 360°, which is the degree measure of a complete revolution about a circle. It is used to show the relative sizes of different categories within a dataset.

Example 10.9 A company's annual budget is divided into four categories: marketing (40%), research and development (30%), operations (20%), and administration (10%) (Figure 10.2).

What Is the Utility of a Pie Graph?

There is an intuitive understanding of pie graphs as a means of expressing the breakdown of a population's properties or characteristics. As an illustration, consider the preference among consumers for particular operating systems among computer users.

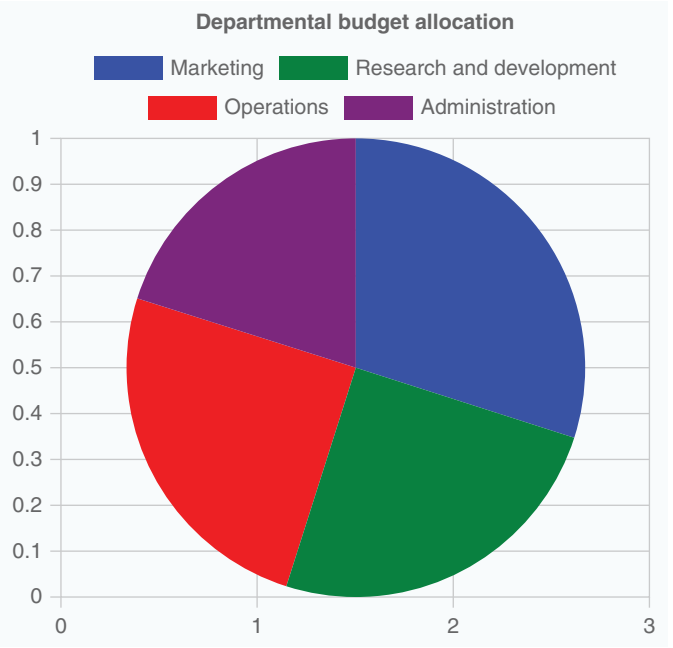
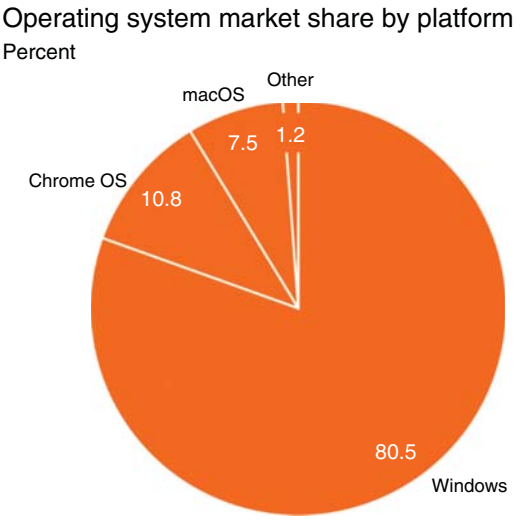


Figure 10.2 Company budget pie chart.



Which operating system is predominantly chosen by computer users? A superficial examination of the image shown above makes evident the dominance of the Windows operating system and the relatively infrequent choice of the macOS. What, in your opinion, is the reason behind the relatively small market share of the macOS, despite the desirability of Apple products among consumers?

10.2.3 Line Graph

A line graph, which consists of a series of data points connected by line segments, shows trends within and changes in data values over the passage of time. These should not be confused with

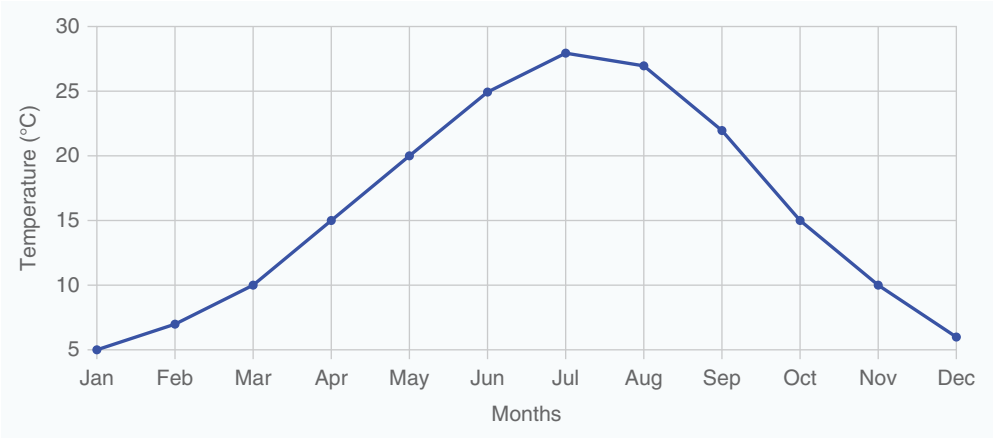
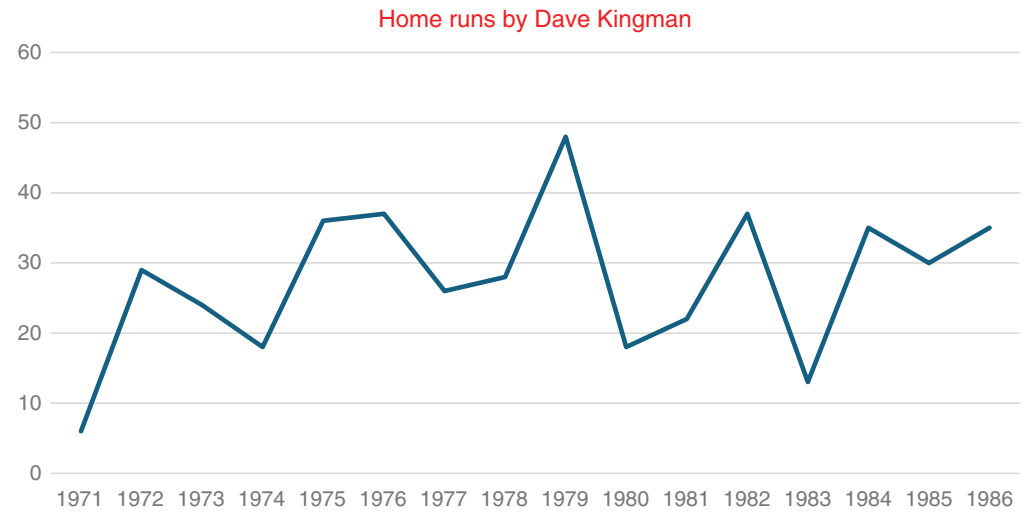


Figure 10.3 Average monthly temperature line graph.

"trendlines," which are used as forecasting tools for a set of data. There is nuance between the two concepts, which we will examine as we proceed.

Example 10.10 The average monthly temperature in a city in the northern hemisphere over a year is as follows: Jan (5 °C), Feb (7 °C), Mar (10 °C), Apr (15 °C), May (20 °C), Jun (25 °C), Jul (28 °C), Aug (27 °C), Sep (22 °C), Oct (15 °C), Nov (10 °C), Dec (6 °C) (Figure 10.3).

Example 10.11 One of the most prolific home run hitters of the 1970s and 1980s was a journeyman outfielder named Dave Kingman. Though he hit 442 home runs during his Major League career, he was not elected to the Hall of Fame. Shown is a line graph of the home runs hit by Dave Kingman, by year, through his career:

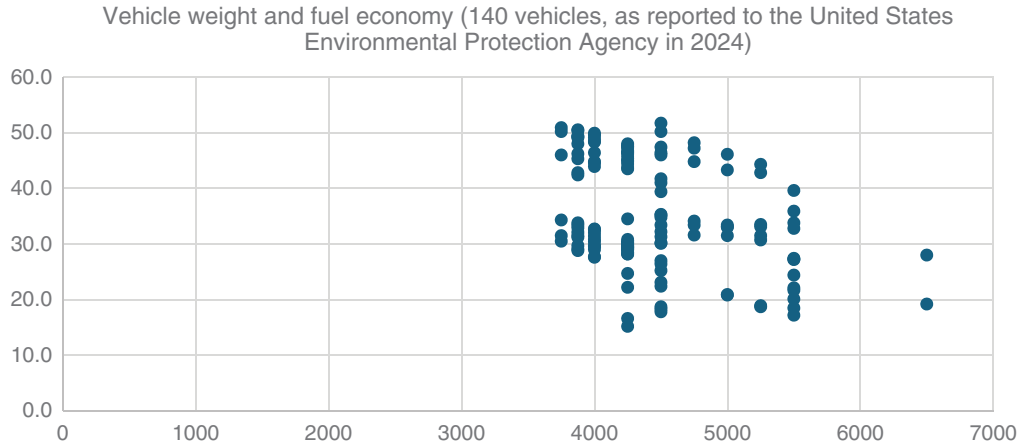


10.2.4 Scatter Plot

A scatter plot is used to show the relationship between two continuous variables. Each data point is represented by a dot on a Cartesian plane, with the x-coordinate representing the independent variable and the y-coordinate representing the dependent variable.

Example 10.12 A researcher collected data on the height (in inches) and weight (in pounds) of 20 individuals. The data points are plotted on a scatter plot to examine the relationship between the two variables (Figure 10.4).

Example 10.13 An automotive magazine wishes to examine the relationship between the weights of vehicles and their fuel economy (miles per gallon). After gathering data from the Environmental Protection Agency on 140 types of vehicles, they found that the data can be represented as:



In the above graph, one can see some observations that are aligned horizontally. This appears to occur most noticeably in the graph, where the weight is between 4000 pounds and 4500 pounds.

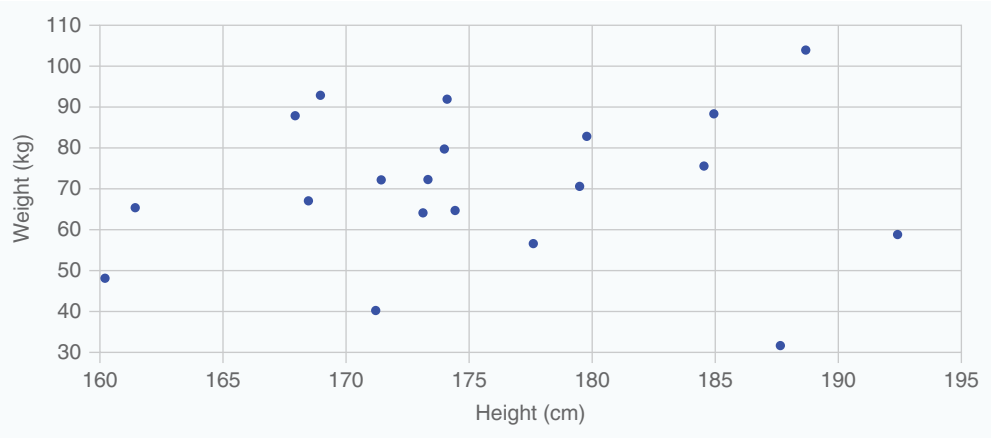


Figure 10.4 Height vs. Weight scatter plot.

Upon superficial examination, what can we conclude from the above graph, and is it surprising? As we move to the right on the graph, the cars are getting heavier. We observe that, as this happens, the miles per gallon for the vehicles decreases. Intuitively, this makes sense: to propel a vehicle that weighs more should require a greater use of fuel, and so the miles per gallon for heavier vehicles should be lower than that for lighter vehicles.

10.2.5 Histogram

A histogram is used to display the distribution of a continuous variable. It consists of a series of adjacent rectangles, where the width of each rectangle represents a class interval, and the height represents the frequency or number of observations within that interval.

Example 10.14 The test scores of 50 students in a class are as follows: 85, 92, 78, 88, 90, 75, 82, 95, 87, 83, 89, 79, 91, 84, 86, 93, 77, 81, 94, 80, 76, 96, 74, 98, 97, 73, 99, 72, 100, 71, 68, 70, 69, 67, 66, 65, 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51 (Figure 10.5).

At first glance, it is common to consider histograms and bar charts to be identical. However, the definition of the two graphs indicates a subtle distinction to be made between the two.

The fundamental difference between histograms and bar graphs will help you to identify the two easily is that there are gaps between bars in a bar graph, while in a histogram, the bars are adjacent to one another. However, there is slightly more to the story; we list some specific differences below:

Bar Chart	Histogram
Represents categorical data	Represents quantitative data
Shows us a comparison of different categories	Shows us the distribution within ranges of a continuous category
Has spaces between the bars	Has no spaces between the bars
All bars are the same width	Bars can be different widths

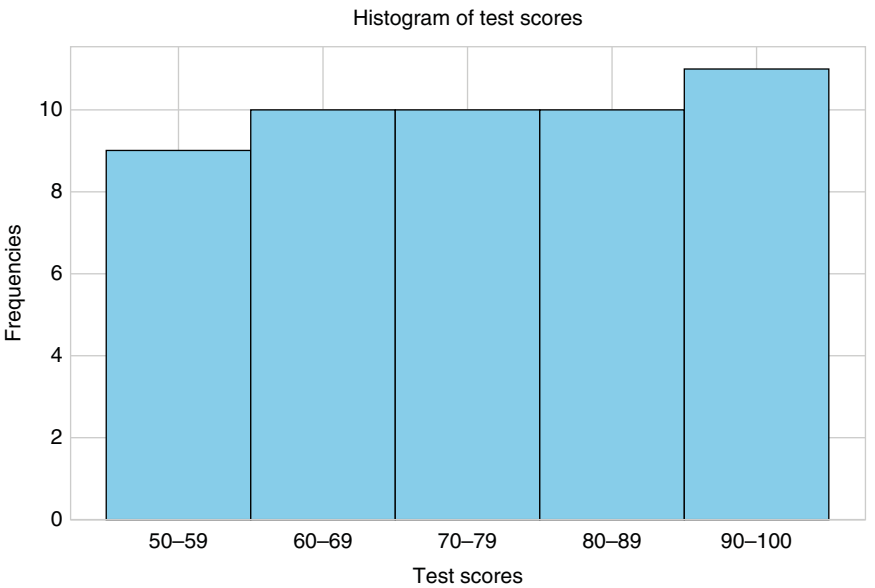


Figure 10.5 Test scores histogram.

10.3 The Measures of Central Tendency

Measures of central tendency are statistical measures that describe the center or middle of a dataset. They provide a single value that represents the typical or central value of the data. The three most common measures of central tendency are the mean, median, and mode. In this section, we will discuss each of these measures in detail and provide examples of their calculations and interpretations.

10.3.1 Mean

The mean, also known as the arithmetic mean or average, is the sum of all values in a dataset divided by the total number of values. It is denoted by the symbol μ (mu) for the population mean and $\bar{x}$ (x-bar) for the sample mean.

Example 10.15 Consider the following dataset of test scores: 78, 85, 92, 88, 90, 75, 82, 95, 87, 83. To calculate the mean, add all the values and divide by the number of values:

$$\begin{aligned}\text{Mean} &= (78 + 85 + 92 + 88 + 90 + 75 + 82 + 95 + 87 + 83) \div 10 \\ &= 855 \div 10 \\ &= 85.5\end{aligned}$$

The mean test score is 85.5.

10.3.2 Median

The median is the middle value in a dataset when the values are arranged in ascending or descending order. If the dataset has an odd number of values, the median is the actual middle value in the ordered dataset. If the dataset has an even number of values, the median is the average of the two middle values in the ordered dataset.

Example 10.16 Determine the median of the following dataset of ages: 25, 32, 18, 47, 56, 39, 42. To find the median, first arrange the values in ascending order:

18, 25, 32, 39, 42, 47, 56.

Since there are 7 values (an odd number), the median is the middle value, which is 39.

Example 10.17 Determine the median of the following dataset: 25, 32, 18, 47, 56, 39, 42, 60. Arranged in ascending order: 18, 25, 32, 39, 42, 47, 56, 60.
The median is the average of the two middle values: $(39 + 42) \div 2 = 40.5$.

10.3.3 Mode

The mode is the value that appears most frequently in a dataset, if such a value exists. A dataset can have no mode (if there is no value which appears with greater frequency than any other value in the dataset), one mode (unimodal, where a single data value has maximal frequency of appearance in the dataset), or multiple modes (bimodal or multimodal, where two or more data values have equal and maximal frequency of appearance in the dataset).

There is potential for conflict with the cases for no mode or multimodal datasets. If each data value in the set appears once, is there no mode or is this an extreme case of multimodal data? We shall agree that multimodal data will only be so regarded if there are multiple appearances of values in the dataset.

Example 10.18 Consider the following dataset of favorite colors, expressed by a group of elementary school students: red, blue, green, blue, yellow, red, blue, orange, blue.

In this dataset, the color "blue" appears four times, which is more than any other color. Therefore, the mode of this dataset is "blue."

Example 10.19 A popular statistic in Major League Baseball is the leading number of home runs in any given season. Considering data of year-by-year leaders (according to http://mlb.com), what is the mode for the set of league-leading home run totals for all seasons from 1901 to the present?

Solution:
The data, on examination, is multimodal. The greatest frequency with which any league-leading home run total occurs is 7; this happens for league-leading home run totals of 32, 44, 48, and 49.

Interpreting Measures of Central Tendency

It is reasonable to ponder which of the measures of central tendency is most accurate in describing the "average" of a dataset. If one is superior to the others, it is natural to question the need for the others.

We find, however, that each of the measures has its own merits and faults, which we should discuss prior to moving forward. To appreciate the distinctions, we first consider the notion of "sensitivity" of a measure. "Sensitivity" refers to how the inclusion of an additional member to a dataset affects the value of the measure.

Consider the following dataset: {2, 2, 3, 5, 6, 6, 6, 7, 10}. What is the mean of this dataset? The sum of the data values is 47, and the mean of the dataset is 5.222, approximately. Now consider the nearly identical dataset: {2, 2, 3, 5, 6, 6, 6, 7, 10, 11}. The sum of the data values is 58, and the mean of the dataset is 5.8. Finally, consider the dataset: {2, 2, 3, 5, 6, 6, 6, 7, 10, 26}. The sum of the data values is 73, and the mean of the dataset is 7.3.

Including a new data value in the second set hardly affects the mean at all; notice it changed from 5.222 to 5.8. The reason the change was so small was that the additional value (11) was not far from the existing values of data in the set. On the other hand, including the new data value in the third set did affect the mean rather significantly. The reason for this is that the new value (26) was rather far from the existing values of the dataset.

A value that lies far from the majority of the data values is called an outlier. The concept "far" is unacceptably vague, though there is an intuitive understanding of what we mean by this notion. We can solidify the idea by defining what we mean by "far," and this can be done using a value called the "Interquartile range." We will not explore this here, however, and will accept the ambiguity of the nonspecific term "far." Referring to the same sets of data, what are the median values of the sets?

{2, 2, 3, 5, 6, 6, 6, 7, 10} has nine data values; this is an odd quantity, and so the middle value of the dataset is the 5th data value, which here is 6.

{2, 2, 3, 5, 6, 6, 6, 7, 10, 11} has 10 data values, and so the middle value of the dataset is the average of the 5th and 6th data values in the set, which is $\frac{6+6}{2} = 6$.

{2, 2, 3, 5, 6, 6, 6, 7, 10, 26} has 10 data values, and so the middle value of the dataset is the average of the 5th and 6th data values in the set, which is $\frac{6+6}{2} = 6$.

Note the inclusion of additional data values in the original set does not affect the value of the median (in this case, there is no affect at all; this is not universally the case). In particular, the inclusion of an outlier does not greatly affect the median.

For the three datasets, the mode is (in each case) 6. The inclusion of an additional data value (in particular, an outlier) does not affect the mode.

Combining these observations, we conclude the following:

- The mean is sensitive to extreme values (outliers) and is best used when the data is not skewed.
- The median is less affected by extreme values and is preferred when the data is skewed or has outliers.
- The mode is useful for categorical data or when identifying the most common value is of interest.

Example 10.20 Calculate the mean, median, and mode of the following dataset representing the number of books read by students in a month: 5, 7, 3, 8, 5, 9, 4, 6, 5, 7.

Mean:

Step 1: Add all the values: $5 + 7 + 3 + 8 + 5 + 9 + 4 + 6 + 5 + 7 = 59$
Step 2: Divide the sum by the number of values: $59 \div 10 = 5.9$

The mean number of books read is 5.9.

Median:

Step 1: Arrange the values in ascending order: 3, 4, 5, 5, 5, 6, 7, 7, 8, 9
Step 2: Identify the middle value: Since there are 10 values (an even number), the median is the average of the two middle values: $(5 + 6) \div 2 = 5.5$

The median number of books read is 5.5.

Mode:

The value that appears most frequently is 5, which occurs three times. Therefore, the mode is 5.

Example 10.21 Calculate the mean, median, and mode of the following dataset representing the heights (in inches) of a group of people: 68, 72, 65, 70, 68, 73, 71, 68, 69, 70, 72, 68.

Mean:

Step 1: Add all the values: $68 + 72 + 65 + 70 + 68 + 73 + 71 + 68 + 69 + 70 + 72 + 68 = 854$
Step 2: Divide the sum by the number of values: $854 \div 12 = 71.17$ (rounded to two decimal places)

The mean height is 71.17 inches.

Median:

Step 1: Arrange the values in ascending order: 65, 68, 68, 68, 68, 69, 70, 70, 71, 72, 72, 73
Step 2: Identify the middle value: Since there are 12 values (an even number), the median is the average of the two middle values: $(69 + 70) \div 2 = 69.5$

The median height is 69.5 inches.

Mode:

The value that appears most frequently is 68, which occurs four times. Therefore, the mode is 68 inches.

Example 10.22 In a small company, the salaries of the employees (in thousands of dollars per year) are given by {23, 28, 28, 31, 33, 33, 33, 33, 92}. What are the values of the mean, median, and mode, and why might each be of use?

Solution:
The mean of the dataset is 37.625, the median is 32, and the mode is 33.

Suppose management and employees had a disagreement over the rate of pay for the employees of this company; which measure of central tendency would be most advantageous for each side of the negotiation? From the perspective of management, the mean would be the most desirable measure to use as an argument against increasing salaries, while from the perspective of labor, either the median or the mode would be preferable.

The objection lodged by labor would be that a particular data value (the salary of $92,000 per year) is distorting the value of the mean, making that measure an unsuitable representation of the average salary of an employee. The position of management would be that the mean is the typically taken view of "average," and thus, the average salary at the company is adequate.

Observe neither side of argument is a lie. At issue is the interpretation of the meaning of the term "average," and it is the resolution of this debate that will decide the outcome of salary negotiations at the company.

10.4 The Measures of Dispersion

Measures of dispersion, also known as measures of variability, describe the spread or variability of a dataset. They provide information about how much the data values deviate from the central tendency measures, such as the mean or median. Understanding the dispersion of a dataset is crucial for making accurate interpretations and comparisons. In this section, we will discuss three common measures of dispersion: range, variance, and standard deviation.

10.4.1 Range

The range is the simplest measure of dispersion and is defined as the difference between the largest and smallest values in a dataset.

Example 10.23 Consider the following dataset of test scores: 78, 85, 92, 88, 90, 75, 82, 95, 87, 83.
To calculate the range:

Largest value: 95
Smallest value: 75

Range = Largest value − Smallest value = 95 – 75 = 20
The range of the test scores is 20.

Example 10.24 Consider the following dataset of salaries (in thousands of dollars per year) for employees at a particular small company: 23, 28, 28, 31, 33, 33, 33, 33, 92.

Solution:

Largest value: 92
Smallest value: 23

Range = Largest value – Smallest value = 92 – 23 = 69

The range of the dataset (in thousands of dollars per year) is 69.

Observe the range is sensitive to the presence of outliers. In the salary example, had the value 92 been omitted from the dataset, the range would have been 33 – 23 = 10. The inclusion of the outlier 92 has changed the range from 10 to 69, which is certainly significant in this context.

10.4.2 Variance

Variance, introduced in Chapter 10, measures the average squared deviation of each data point from the mean. It quantifies the spread of the data relative to the mean. The population variance is denoted by σ^2 (sigma squared), and the sample variance is denoted by s^2.

The formulas for variance are as follows:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n}$$

and

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

Observe the formula for a population is not the same as the formula for a sample. The change from n to $(n - 1)$ in the denominator is referred to as "Bessel's Correction," named for Friedrich Wilhelm Bessel (1784–1845), a mathematician from Germany.

A full explanation for the correction lies beyond the scope of our discussion, but suffice it to say the modification corrects for the bias in the estimation of the population variance and partially corrects the bias in the estimation of the population standard deviation. We refer you to a full course in statistics for a more in-depth analysis of the need for and proof of Bessel's Correction.

When we use the term "variance," it is to be understood we mean the population variance, rather than the sample variance. If we intend to compute the sample variance, we shall explicitly say so.

Example 10.25 Calculate the variance of the following sample taken from the dataset: 4, 7, 3, 8, 5.

Step 1: Calculate the mean: $(4 + 7 + 3 + 8 + 5) \div 5 = 5.4$

Step 2: Subtract the mean from each value and square the differences:

$(4 - 5.4)^2 = 1.96$
$(7 - 5.4)^2 = 2.56$
$(3 - 5.4)^2 = 5.76$
$(8 - 5.4)^2 = 6.76$
$(5 - 5.4)^2 = 0.16$

Step 3: Add the squared differences: $1.96 + 2.56 + 5.76 + 6.76 + 0.16 = 17.2$

Step 4: Divide the sum by the number of values in the dataset: $17.2 \div (5ca) = 4.3$

The variance is 3.44.

Example 10.26 Calculate the variance of the following sample taken from a larger dataset: 4, 7, 3, 8, 5.

Step 1: Calculate the mean: $(4 + 7 + 3 + 8 + 5) \div 5 = 5.4$

Step 2: Subtract the mean from each value and square the differences:

$(4 - 5.4)^2 = 1.96$
$(7 - 5.4)^2 = 2.56$

$(3 - 5.4)^2 = 5.76$
$(8 - 5.4)^2 = 6.76$
$(5 - 5.4)^2 = 0.16$

Step 3: Add the squared differences: $1.96 + 2.56 + 5.76 + 6.76 + 0.16 = 17.2$
Step 4: Divide the sum by the number of values minus 1 (for sample variance): $17.2 \div (5 - 1) = 4.3$
The sample variance is 4.3.

Observe the sample variance is larger than that of the population. This is caused, computationally, by the change in the denominator from n to $(n - 1)$. When the size of a denominator decreases, the value of a fraction increases. A partial explanation for this is that when you compute some property of a sample, that property will inevitably vary about its population value: sometimes above and sometimes below. Moreover, provided the population property is well-defined and finite, the amount of that variation is smaller with larger samples.

10.4.3 Standard Deviation

The standard deviation is the square root of the variance. It measures the average distance of each data point from the mean and is expressed in the same units as those used in the original dataset. The population standard deviation is denoted by σ (sigma), and the sample standard deviation by s.

Intuitively, the meaning of "standard deviation" should appear reasonable: "standard" refers to "typical," and "deviation" to "difference from." If we put the two concepts together, we get a general understanding of what is meant by the term "standard deviation."

Example 10.27 Calculate the standard deviation of the dataset sample from the variance example.

Given: Sample variance = 4.3
To find the sample standard deviation, take the square root of the sample variance:
Sample standard deviation = $\sqrt{(\text{sample variance})} = \sqrt{4.3} \approx 2.07$
The sample standard deviation is approximately 2.07.

Example 10.28 The test scores of students in a class on Discrete Mathematics are {81, 83, 68, 77, 75, 75, 91, 59, 80, 83, 71, 72, 77, 74}. What is the standard deviation of the population of these students?

Solution:
Using the formula for variance of a population (recalling here the denominator is "n"; do you understand why?), and computing the mean as approximately 76.143, we find the population variance is approximately 54.6939 points^2. We have no ready understanding of what "points squared" would represent, so this is not terribly informative.

We take the square root of the variance and find the standard deviation to be 7.3955, approximately. We shall interpret this value in what follows.

Interpreting Measures of Dispersion

- A larger range, variance, or standard deviation indicates that the data points are more spread out from the mean, while smaller values suggest that the data points are more tightly clustered around the mean.

- The standard deviation is particularly useful because it is expressed in the same units as the original data, making it easier to interpret and compare across different datasets (the same can be said of the range, regarding the units of measure being the same as those of the dataset, but the range is extremely sensitive to outliers).

Example 10.29 Calculate and interpret the range, variance, and standard deviation of the following sample dataset representing the ages of a group of people: 25, 32, 18, 47, 56, 39, 42.

Range:

Largest value: 56
Smallest value: 18

Range = Largest value − Smallest value = 56 − 18 = 38

The span of values for the ages of the individuals in this group is 38 years. That is, the oldest person is 38 years older than the youngest person.

Variance:

Step 1: Calculate the mean: $(25 + 32 + 18 + 47 + 56 + 39 + 42) \div 7 = 37$

Step 2: Subtract the mean from each value and square that difference:

$(25 - 37)^2 = 144$
$(32 - 37)^2 = 25$
$(18 - 37)^2 = 361$
$(47 - 37)^2 = 100$
$(56 - 37)^2 = 361$
$(39 - 37)^2 = 4$
$(42 - 37)^2 = 25$

Step 3: Add the squared differences: $144 + 25 + 361 + 100 + 361 + 4 + 25 = 1020$

Step 4: Divide the sum by the number of values minus 1 (for sample variance): $1020 \div (7 - 1) = 170$

The variance is 170 years^2.

Because it uses squared units rather than the natural data units, the interpretation is less intuitive. Higher values indicate greater variability, but there is no intuitive interpretation for specific values. Since the value 170 is difficult to interpret as being "greater" or "lesser" because we are looking only at a single dataset, we cannot effectively interpret what this value tells us. Let us move on to the standard deviation.

Standard Deviation:

$$\text{Sample standard deviation} = \sqrt{(\text{sample variance})} = \sqrt{170} \approx 13.04 \text{ years.}$$

The standard deviation tells us the typical difference of a given data value from the mean. In this case, the value of 13.04 years tells us that, in this dataset, the typical difference from the mean of 37 years old for an individual is 13.04 years.

As a frame of reference, we might expect (as is normally the case – keep that term in mind) that most individuals in this dataset have ages that are between 24 and 50 years, approximately. Let us order the dataset for this group and see if this is reasonable.

If we examine the data, it turns out five of the seven data values are within that range.

Example 10.30 Calculate and interpret the range, variance, and standard deviation of the following dataset, representing the number of hours a sample of students in a given class spent studying for an exam: 3, 6, 4, 8, 5, 7, 6, 4.

Range:

Largest value: 8
Smallest value: 3

Range = Largest value − Smallest value = 8 – 3 = 5

The number of hours spent studying for the exam among the students in this sample ranges from 3 to 8 hours.

Variance:

Step 1: Calculate the mean: $(3+6+4+8+5+7+6+4) \div 8 = 5.375$

Step 2: Subtract the mean from each value and square the differences:

$(3 - 5.375)^2 = 5.64$
$(6 - 5.375)^2 = 0.39$
$(4 - 5.375)^2 = 1.89$
$(8 - 5.375)^2 = 6.89$
$(5 - 5.375)^2 = 0.14$
$(7 - 5.375)^2 = 2.64$
$(6 - 5.375)^2 = 0.39$
$(4 - 5.375)^2 = 1.89$

Step 3: Add the squared differences: $5.64 + 0.39 + 1.89 + 6.89 + 0.14 + 2.64 + 0.39 + 1.89 = 19.87$

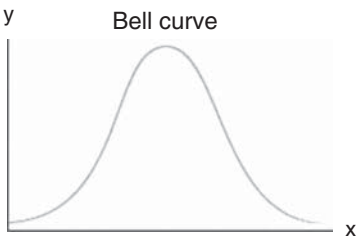
Step 4: Divide the sum by the number of values minus 1 (for sample variance):
$19.87 \div (8-1) \approx 2.84$

Standard Deviation:

Sample standard deviation = $\sqrt{(\text{sample variance})} = \sqrt{2.84} \approx 1.68$

10.5 The Normal Distribution

The normal distribution, also known as the Gaussian distribution, or the "bell curve," is a continuous probability distribution that is widely used in statistics. It is symmetric about its mean and has a hump-shaped curve reminiscent of the outline of a bell.



Many natural phenomena and datasets follow a normal distribution, making it an essential concept in statistical analysis. It is the fact that data typically follows this distribution that prompts us to refer to the distribution as "normal."

In this section, we will discuss the properties of the normal distribution, its probability density function, and the **standard** normal distribution.

Properties of the Normal Distribution

- Symmetry: The normal distribution is symmetric about its mean, meaning that the left and right halves of the curve are mirror images of each other. This can be seen in the image provided above.
- Mean, median, and mode: In a normal distribution, all three are equal and located at the center of the distribution.
- Skewness: The normal distribution has no skewness, which means that it is not stretched or tilted to either side.

Probability Density Function

The probability density function (PDF) of a normal distribution is given by:

$$f(x) = (1 / (\sigma\sqrt{[2\pi]})) * e\text{^}(-((x - \mu)^2 / (2\sigma^2)))$$

where

- μ is the mean of the distribution
- σ is the standard deviation of the distribution
- π is the mathematical constant pi (approximately 3.14159)
- e is the mathematical constant (approximately 2.71828)

The PDF represents the relative likelihood of a random variable taking on a specific value within the distribution. In the particular case of the standard normal distribution, the mean is zero, and the standard deviation is one.

10.5.1 Standard Normal Distribution

The **standard normal distribution** is a special case of the normal distribution, with a mean of 0 and a standard deviation of 1. It is denoted by $Z \sim N(0, 1)$. Any normal distribution can be converted to a standard normal distribution using the z-score formula; we call this "standardization":

$$Z = (X - \mu)/\sigma$$

where

- X is a value from the original normal distribution
- μ is the mean of the original normal distribution
- σ is the standard deviation of the original normal distribution
- z-scores are numerical only; they have no units of measure

- z-scores can be positive or negative; a negative z-score indicates the data value is less than the mean, a positive z-score indicates the data value is more than the mean, a z-score of zero indicates the data value is equal to the mean;
- larger z-scores (in absolute value) indicate data values are further from the mean

Why do we concern ourselves with a "standard" normal distribution? In the days prior to the advent of computers, it was inconvenient to deal with normal distributions on a case-by-case basis. It proved convenient to "standardize" the normal distribution in such a manner that a single table of values could be used to analyze data that was normally distributed. Though technology could easily make this reference to a standardized version of the distribution, tradition compels us to do so.

As we shall see, the use of standardization of data values to z-scores allows us to draw statistical conclusions about a data value's "place" within a dataset. First, we shall calculate a z-score so the method is clear and then shall move to the interpretation.

Example 10.31 In the United States of America, the average height of a man is approximately 69 inches, with a standard deviation of 2.3 inches. What is the z-score for a man whose height is 76 inches (such as your author!)?

Solution:
The z-score is $\frac{x-\mu}{\sigma} = \frac{76-69}{2.3} \approx 3.04$

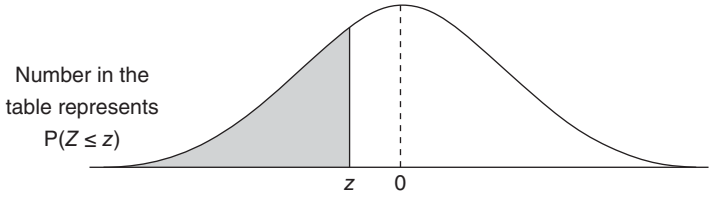
The standard normal distribution has been the subject of extensive research, and much work has been devoted to the creation of "standard normal distribution tables." The purpose of these tables is to allow us to establish percentages within which potential values of a dataset will fall when examining datasets.

Computers have, to a degree, made the use of tables redundant, but the analysis of these tables is instructive and will be examined here.

Shown are two pages of standard normal distribution values; they may (at first) seem intimidating but be assured their use is not complicated. To use the tables, we must view z-scores as being rounded to two decimal places and decomposed into two parts (the whole number and tenth's part and the hundredth's part).

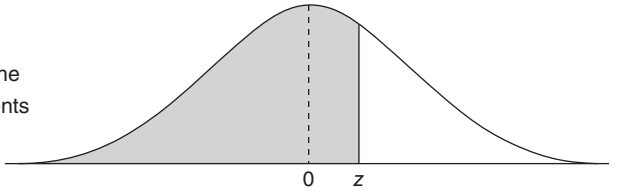
That is, a z-score of 2.18 should be considered as $2.1 + 0.08$. Other illustrations are:

z-score	decomposition
1.15	$1.1 + 0.05$
-2.08	$-2.0 + (-0.08)$
0.06	$0.0 + 0.06$
-1.07	$-1.0 + (-0.07)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
–3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Number in the table represents $P(Z \le z)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

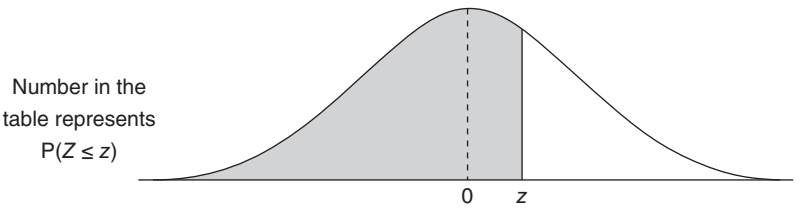
How do we use the z-score to find a particular table entry? After decomposing the z-score into its constituent parts, we look across the row beginning with the first two digits of the z-score and down the column corresponding to the final digit of the z-score, and this gives us a table entry. We now do so with the decompositions mentioned a moment ago:

z-score	decomposition	table value
1.15	$1.1 + 0.05$	0.8749
−2.08	$-2.0 + (-0.08)$	0.0188
0.06	$0.0 + 0.06$	0.5239
−1.07	$-1.0 + (-0.07)$	0.1423

If the z-score exceeds the maximum values listed (less than −3.69 or more than 3.69), we can use the value of 0.0001 and 0.9999, respectively.

The table values are percentiles. They refer to the percentage of the population whose value is less than or equal to the data value for which we found the z-score. If we look at (1 – table entry), we get the percent of the population that is greater than our data value.

The bell curve figure above each of the halves of the table offers an important visualization of this concept of probability.



We have mentioned the probability of all possible outcomes (combined) is 1.00 or 100%. The probability corresponds to the area beneath this probability distribution graph; by this, we mean the area beneath the standard normal distribution curve's bell graph is exactly 1 square unit, corresponding to probability 1.00.

Once a z-score for the standard normal distribution has been computed and the corresponding table entry found, the entry can be visualized as the area beneath the standard normal distribution curve and to the left of the z-score on the horizontal axis.

A useful adjunct to this result can be gleaned if we consider an illustration of a commonsense notion we can co-opt for the purpose of shortening our path to the answer of a statistical question related to the normal distribution.

Suppose it is known that 40 vehicles in a commercial fleet currently have less than 100,000 miles on their odometer and that 18 vehicles in that same fleet have less than 60,000 miles on their odometer. A moment's reflection tells us that 22 of the vehicles in the fleet have between 60,000 and 100,000 miles on their odometer. This observation can be extended to our newly discovered use of a standard normal distribution table.

If we wish to find the percentage of the members of a population whose values fall between two extremes, we can find the z-score of each of the two extremes, locate the corresponding table values, and then calculate their difference. This notion aligns precisely with the illustration regarding the mileage registered on the odometers of the fleet of commercial vehicles stated above.

Example 10.32 In the United States of America, the average height of a man is approximately 69 inches, with a standard deviation of 2.3 inches. What is the z-score for a man whose height is 76 inches (such as your author!)? How do we interpret this z-score, using the standard normal distribution table?

Solution:
Recall the z-score is $\frac{x-\mu}{\sigma} = \frac{76-69}{2.3} \approx 3.04$; the corresponding table value is 0.9988; this tells us 99.88% of the population of men in the United States have a height that is less than or equal to the height of your instructor.

This is, in a sense, astounding. It reveals us the power of detecting a normal distribution of data, an illustration of which is the notion of "stocking" clothing of a particular size. Since your professor is taller than 99.88% of all men in the United States, we can see that a clothing manufacturer would be ill-advised to keep on hand great quantities of an article of clothing designed for such a tall individual.

Example 10.33 Heights of women in the United States are normally distributed. Caitlin Clark, basketball sensation of the WNBA (Women's National Basketball Association), is 6 feet tall (72 inches). The average height of a woman in the United States is 63.5 inches, with a standard deviation of 2.3 inches (yes, this is the same as the standard deviation for men—it is a coincidence, perhaps, but true). What percentage of women in the United States are shorter than Ms. Clark?

Solution:
The z-score for Ms. Clark's height is:

$$z = \frac{72 - 63.5}{2.3} = \frac{8.5}{2.3} \approx 3.70$$

Since this exceeds the largest z-score on the table, we use the entry 0.9999 and conclude that Ms. Clark is taller than 99.99% of all women in the United States.

Example 10.34 The average height of a rookie in the WNBA is normally distributed, with a mean of 73 inches and a standard deviation of 3 inches. What percentage of WNBA rookies are shorter than Caitlin Clark?

Solution:

$$z = \frac{72 - 73}{3} \approx -0.33$$

The corresponding table entry for this z-score is 0.3669; from this, we conclude Ms. Clark is taller than 36.69% of rookie players in the WNBA.

In a way this is surprising but, when we consider how much taller WNBA players are than the typical woman in the United States, the result seems eminently reasonable. It is important to consider context when assessing the results of a normal distribution problem.

Example 10.35 Suppose the heights of adult males in a population are normally distributed with a mean of 70 inches and a standard deviation of 2.3 inches. What is the probability that a randomly selected adult male has a height between 67 and 73 inches?

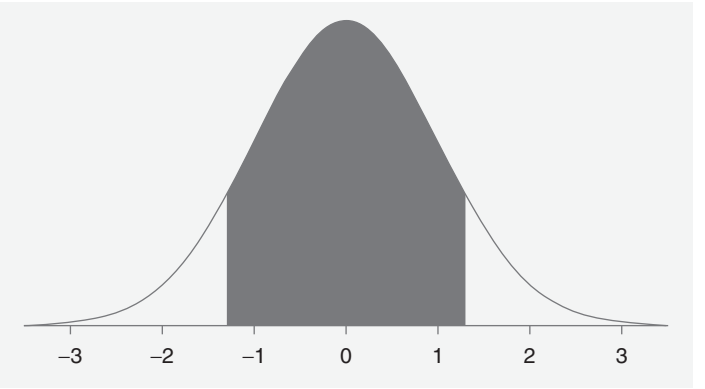
Step 1: Convert the given heights to z-scores using the z-score formula:

For 67 inches: $Z1 = (67 - 70)/2.3 = -1.30$
For 73 inches: $Z2 = (73 - 70)/2.3 = 1.30$

Step 2: Find the area under the standard normal curve between Z1 and Z2 using a Z-table or a calculator:

$$P(-1.30 < Z < 1.30) \approx 0.9032 - 0.0968 = 0.8064$$

Therefore, the probability that a randomly selected adult male has a height between 67 and 73 inches is approximately 0.8064 or 80.64%.



Example 10.36 The weights of apples harvested from a particular orchard are normally distributed with a mean of 150 g and a standard deviation of 20 g. What is the probability that a randomly selected apple weighs less than 130 g?

Step 1: Convert 130 g to a z-score using the z-score formula:

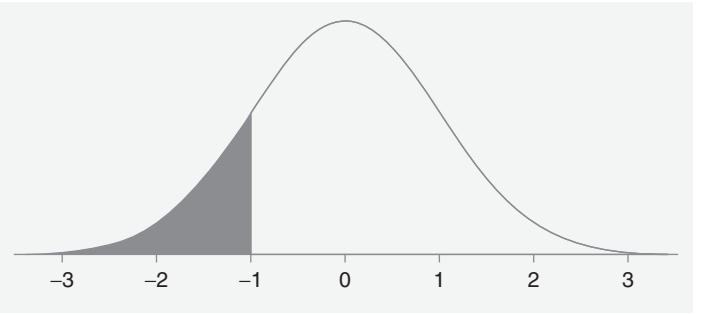
$$Z = (130–150)\,/20 = -1$$

Step 2: Find the area to the left of $Z = -1$ using a Z-table or a calculator:

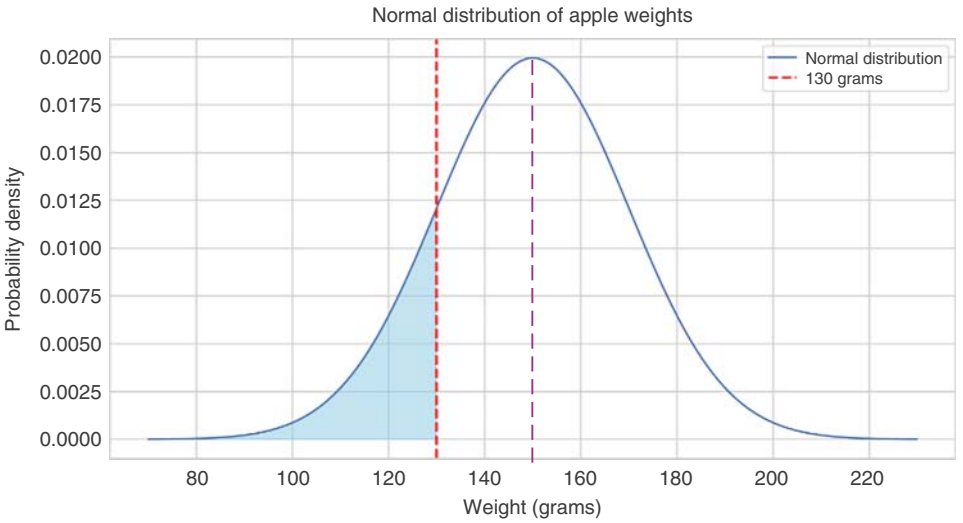
$$P(Z < -1) \approx 0.1587$$

Therefore, the probability that a randomly selected apple weighs less than 130 g is approximately 0.1587 or 15.87%.

Visually:



At this juncture, we point out that the use of a standard normal distribution graph can be modified to be the graph of the normal distribution of the data itself. That is to say, rather than showing the standardized mean and standard deviation values of 0 and 1, we could relabel the horizontal axis to show the units native to the problem. Here, the mean would be 150, and the standard deviation would be 20; this would modify the graph to be:



Observe that this perspective allows us to see the result using the original conceptual units from the problem being addressed. Computationally, nothing is gained by doing so. It is, rather, a matter of esthetics, offering potential clarification for the casual interpreter of the result.

Example 10.37 The IQ scores of a population are normally distributed with a mean of 100 and a standard deviation of 15. What is the probability that a randomly selected individual has an IQ score between 115 and 130?

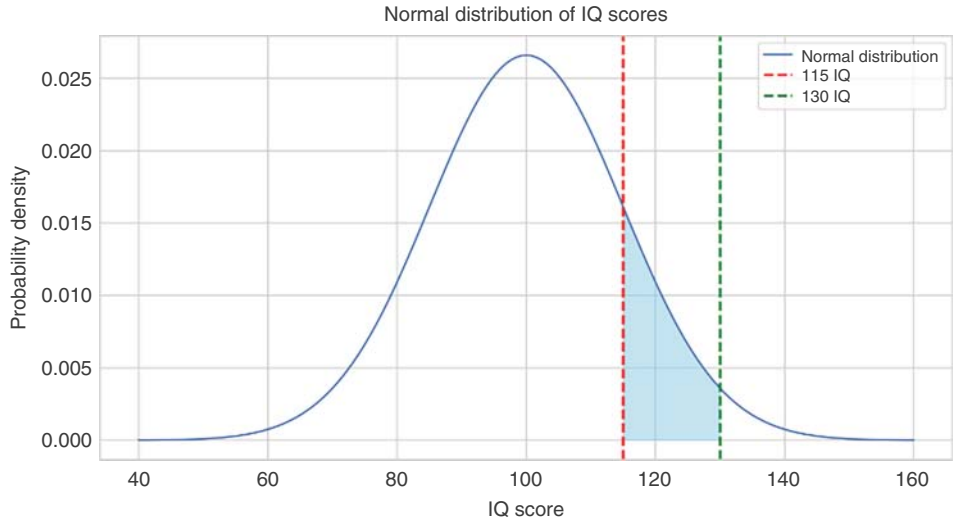
Step 1: Convert the given IQ scores to z-scores using the z-score formula:

For 115: $Z1 = (115–100)/15 = 1$
For 130: $Z2 = (130–100)/15 = 2$

Step 2: Find the area under the standard normal curve between $Z1$ and $Z2$ using a Z-table or a calculator:

$$P(1 < Z < 2) \approx 0.1359$$

Therefore, the probability that a randomly selected individual has an IQ score between 115 and 130 is approximately 0.1359 or 13.59%.



Example 10.38 In naval parlance, a "corvette" is a fast escort vessel displacing 500–700 tons, typically used as part of a convoy. Suppose the navy of a particular country is known to have 72 corvettes currently in service and that the age of these vessels is normally distributed with a mean of 25.8 years and a standard deviation of 13.4 years. What percentage of the navy's corvettes are less than four-years old?

Solution:
The z-score for age four years is $z = \frac{4-25.8}{13.4} = \frac{-21.8}{13.4} \approx -1.63$. Using the standard normal distribution chart, we find the corresponding table entry is 0.0516, and therefore 5.16% of the corvettes in the navy are less than four-years old. Such an observation could be used to argue a modernization program for the navy could be in order.



Example 10.39 The heights of sunflowers grown in a garden are normally distributed with a mean of 60 cm and a standard deviation of 5 cm. If a sunflower is randomly selected, what is the probability that its height is greater than 65 cm?

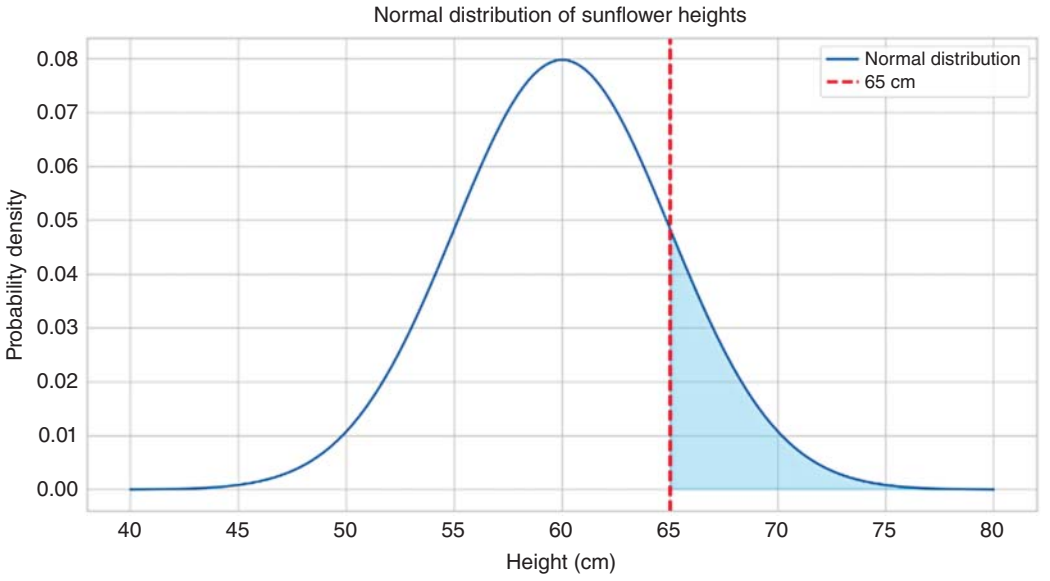
Step 1: Convert 65 cm to a z-score using the z-score formula:

$$Z = (65–60) / 5 = 1$$

Step 2: Find the area to the right of $Z = 1$ using a Z-table or a calculator:

$$P(Z > 1) \approx 0.1587$$

Therefore, the probability that a randomly selected sunflower has a height greater than 65 cm is approximately 0.1587 or 15.87%.



Example 10.40 The battery life of a certain brand of smartphones is normally distributed with a mean of 20 hours and a standard deviation of two hours. If a phone is randomly selected, what is the probability that its battery life is between 18 and 22 hours?

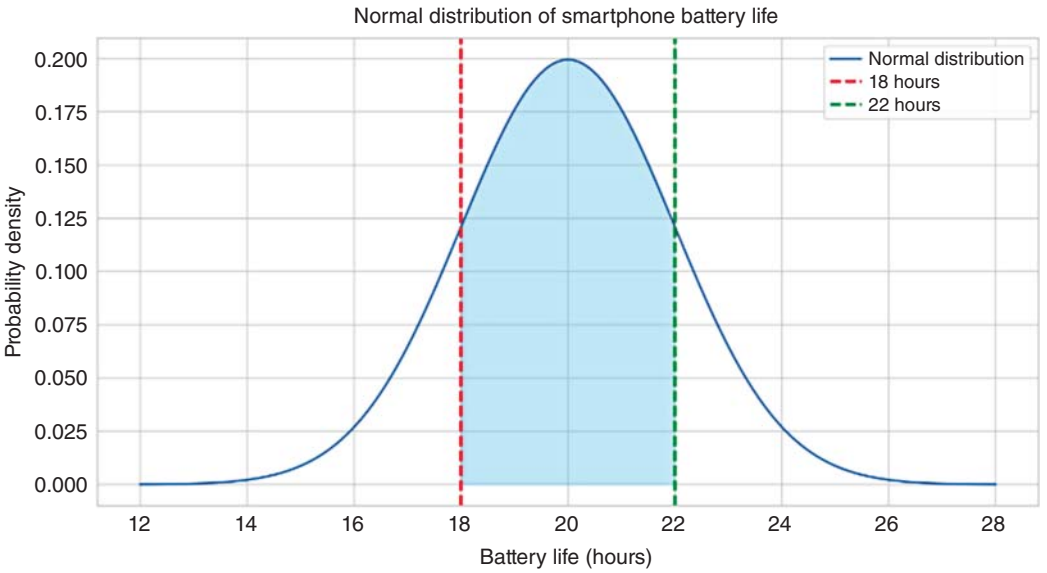
Step 1: Convert the given battery life values to z-scores using the z-score formula:

For 18 hours: $Z1 = (18–20) / 2 = -1$
For 22 hours: $Z2 = (22–20) / 2 = 1$

Step 2: Find the area under the standard normal curve between $Z1$ and $Z2$ using a Z-table or a calculator:

$$P(-1 < Z < 1) \approx 0.6826$$

Therefore, the probability that a randomly selected phone has a battery life between 18 and 22 hours is approximately 0.6826 or 68.26%.

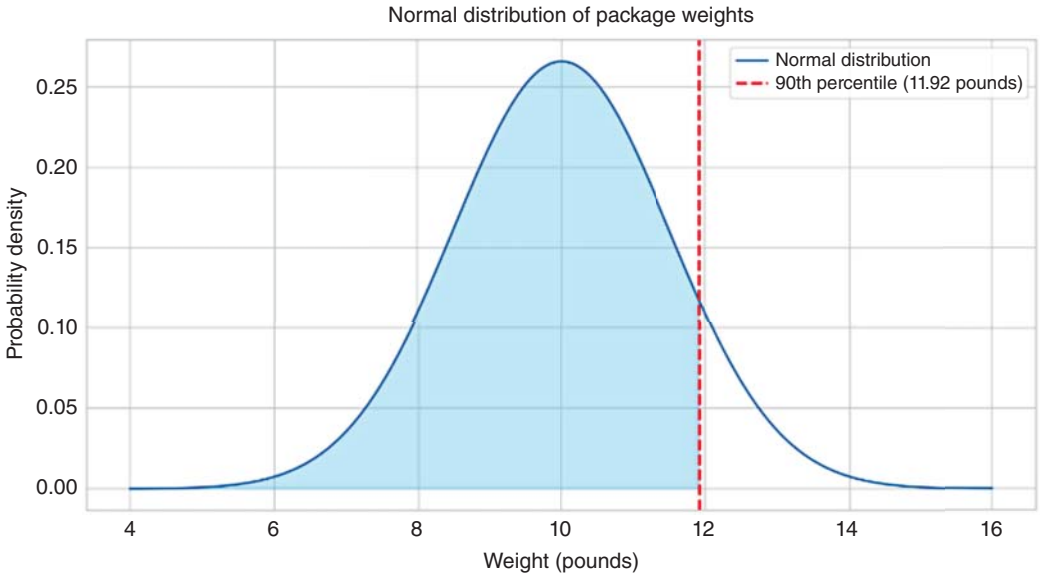


Example 10.41 The weights of packages shipped by a company are normally distributed with a mean of 10 pounds and a standard deviation of 1.5 pounds. What weight represents the 90th percentile of the distribution?

Step 1: Identify the z-score corresponding to the 90th percentile using a Z-table or a calculator:
For the 90th percentile, $Z \approx 1.28$
Step 2: Use the z-score formula to find the corresponding weight:

$$X = \mu + (Z \times \sigma)$$
$$X = 10 + (1.28 \times 1.5) = 11.92 \text{ pounds}$$

Therefore, a weight of approximately 11.92 pounds represents the 90th percentile of the distribution.



Example 10.42 The diameters of a certain type of machine part are normally distributed with a mean of 5 cm and a standard deviation of 0.2 cm. If the acceptable range for the part diameter is between 4.7 cm and 5.3 cm, what percentage of the parts will be within this range?

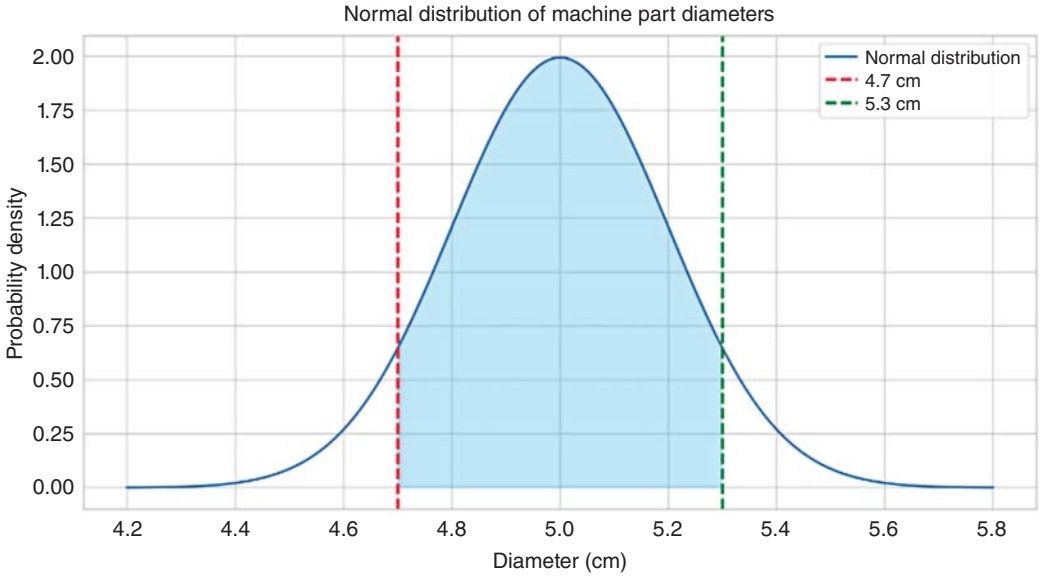
Step 1: Convert the given diameters to z-scores using the z-score formula:

For 4.7 cm: $Z1 = (4.7–5) / 0.2 = -1.5$
For 5.3 cm: $Z2 = (5.3–5) / 0.2 = 1.5$

Step 2: Find the area under the standard normal curve between Z1 and Z2 using a Z-table or a calculator:

$$P(-1.5 < Z < 1.5) \approx 0.8664$$

Therefore, approximately 86.64% of the machine parts will have diameters within the acceptable range of 4.7–5.3 cm.



10.6 The Binomial Distribution

There are three characteristics of a **binomial experiment**:

1) There are a fixed number of trials. The letter n denotes the number of trials.
2) There are only two possible outcomes, called *success* and *failure*, for each trial. The outcome that we are measuring is defined as a *success*, while the other outcome is defined as a *failure*. The letter p denotes the probability of a success on one trial, and q denotes the probability of a failure on one trial. $p + q = 1$.
3) The n trials are independent and are repeated using identical conditions. Because the n trials are independent, the outcome of one trial does not help in predicting the outcome of another trial.
4) The probability of success, denoted by p, remains constant for each trial.

The binomial distribution is a discrete probability distribution that describes the number of successes in a binomial experiment. This distribution is an important concept in statistics and has numerous applications in fields such as biology, engineering, and economics. In this section, we will discuss the characteristics of the binomial distribution, its probability mass function, and provide examples to illustrate its use.

Probability Mass Function

The probability mass function (PMF) of a binomial distribution is given by:

$$\mathrm{P}(X = k) = \mathrm{C}(n, k) \times p\text{^}k \times (1 - \mathrm{p})\text{^}(n - k)$$

where

- X is the random variable representing the number of successes
- n is the total number of trials
- k is the number of successes ($0 \leq k \leq n$)
- p is the probability of success in a single trial
- $\mathrm{C}(n, k)$ is the binomial coefficient, calculated as $n!$ / $(k! \times (n - k)!)$

Example 10.43 A fair coin is tossed 10 times. What is the probability of getting exactly six heads?

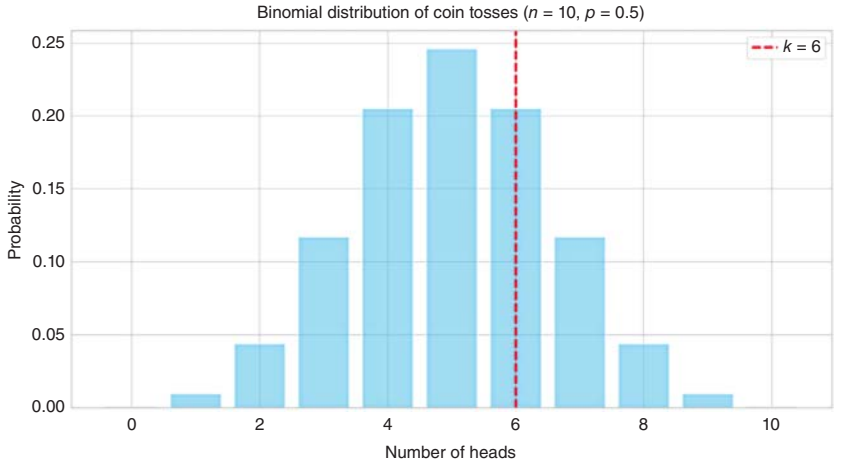
Solution:

$n = 10$ (total number of trials)
$k = 6$ (number of successes, i.e., heads)
$p = 0.5$ (probability of success, i.e., getting a head)

Using the PMF formula:

$$\begin{aligned}\mathrm{P}(X = 6) &= \mathrm{C}(10, 6) \times 0.5\text{^}6 \times (1 - 0.5)\text{^}(10 - 6) \\ &= (10!/[6! \times 4!]) \times 0.5\text{^}6 \times 0.5\text{^}4 \\ &= 210 \times 0.015625 \times 0.0625 \\ &\approx 0.2051\end{aligned}$$

Therefore, the probability of getting exactly 6 heads in 10 tosses of a fair coin is approximately 0.2051 or 20.51%.



Example 10.44 A machine produces defective items with a probability of 0.05. If 20 items are randomly selected, what is the probability that at most 2 items are defective?

Given:

- $n = 20$ (total number of trials)
- $p = 0.05$ (probability of success, i.e., a defective item)

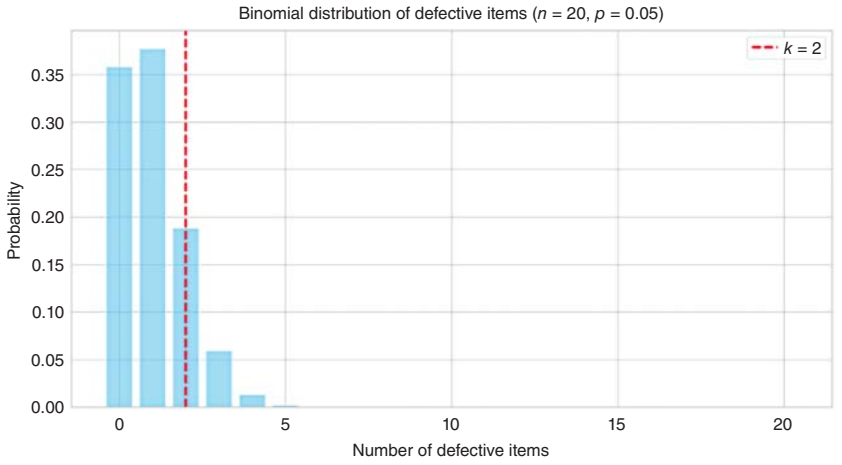
To find the probability of at most two defective items, we need to calculate the sum of probabilities for 0, 1, and 2 defective items:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

Using the PMF formula for each term:

$$P(X = 0) = C(20, 0) \times 0.05\text{^}0 \times (1–0.05)\text{^}(20–0) \approx 0.3585$$
$$P(X = 1) = C(20, 1) \times 0.05\text{^}1 \times (1–0.05)\text{^}(20–1) \approx 0.3774$$
$$P(X = 2) = C(20, 2) \times 0.05\text{^}2 \times (1–0.05)\text{^}(20–2) \approx 0.1887$$
$$P(X \leq 2) = 0.3585 + 0.3774 + 0.1887 \approx 0.9246$$

Therefore, the probability that at most 2 out of 20 randomly selected items are defective is approximately 0.9246 or 92.46%.



Example 10.45 A basketball player has a free-throw success rate of 80%. If the player attempts 15 free throws, what is the probability of making at least 10 of them?

Given:

- $n = 15$ (total number of trials)
- $p = 0.8$ (probability of success, i.e., making a free throw)

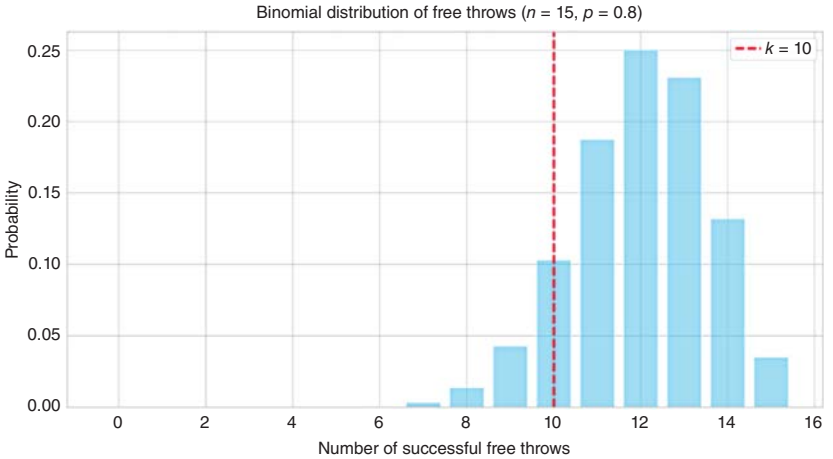
To find the probability of making at least 10 free throws, we need to calculate the sum of probabilities for 10, 11, 12, 13, 14, and 15 successful attempts:

$$P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) + P(X = 15)$$

Using the PMF formula for each term:

$$P(X = 10) = C(15, 10) \times 0.8\text{^}10 \times (1–0.8)\text{^}(15–10) \approx 0.103$$
$$P(X = 11) = C(15, 11) \times 0.8\text{^}11 \times (1–0.8)\text{^}(15–11) \approx 0.188$$
$$P(X = 12) = C(15, 12) \times 0.8\text{^}12 \times (1–0.8)\text{^}(15–12) \approx 0.250$$
$$P(X = 13) = C(15, 13) \times 0.8\text{^}13 \times (1–0.8)\text{^}(15–13) \approx 0.231$$
$$P(X = 14) = C(15, 14) \times 0.8\text{^}14 \times (1–0.8)\text{^}(15–14) \approx 0.132$$
$$P(X = 15) = C(15, 15) \times 0.8\text{^}15 \times (1–0.8)\text{^}(15–15) \approx 0.035$$

$P(X \geq 10)$ is the sum of these probabilities, and so we conclude the probability of making at least 10 out of 15 free throws is approximately 0.939 or 93.9%.



Example 10.46 A company manufactures light bulbs, and the probability of a bulb being defective is 0.02. If a batch of 50 bulbs is selected, what is the probability that exactly 3 bulbs are defective?

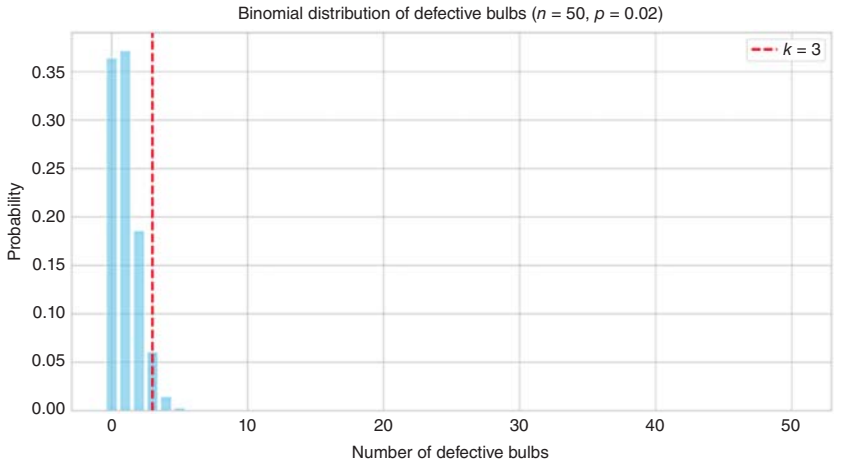
Given:

$n = 50$ (total number of trials)
$k = 3$ (number of successes, i.e., defective bulbs)
$p = 0.02$ (probability of success, i.e., a defective bulb)

Using the PMF formula:

$$P(X = 3) = C(50, 3) \times 0.02\text{^}3 \times (1 - 0.02)\text{^}(50 - 3)$$
$$= (50!/[3! \times 47!]) \times 0.02\text{^}3 \times 0.98\text{^}47$$
$$\approx 0.0607$$

Therefore, the probability of having exactly 3 defective bulbs in a batch of 50 is approximately 0.0607 or 6.07%.



Example 10.47 A survey found that 60% of people prefer a particular brand of chocolate. If 20 people are randomly selected, what is the probability that fewer than eight people prefer this brand?

Given:

- $n = 20$ (total number of trials)
- $p = 0.6$ (probability of success, i.e., preferring the brand)

To find the probability that fewer than 8 people prefer the brand, we need to calculate the sum of probabilities for 0, 1, 2, 3, 4, 5, 6, and 7 people preferring the brand:

$$P(X < 8) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$$

Using the PMF formula for each term:

$P(X = 0) = C(20, 0) \times 0.6\text{^}0 \times (1–0.6)\text{^}(20–0) \approx 0.0000$
$P(X = 1) = C(20, 1) \times 0.6\text{^}1 \times (1–0.6)\text{^}(20–1) \approx 0.0000$
$P(X = 2) = C(20, 2) \times 0.6\text{^}2 \times (1–0.6)\text{^}(20–2) \approx 0.0000$
$P(X = 3) = C(20, 3) \times 0.6\text{^}3 \times (1–0.6)\text{^}(20–3) \approx 0.0000$
$P(X = 4) = C(20, 4) \times 0.6\text{^}4 \times (1–0.6)\text{^}(20–4) \approx 0.0003$
$P(X = 5) = C(20, 5) \times 0.6\text{^}5 \times (1–0.6)\text{^}(20–5) \approx 0.0013$
$P(X = 6) = C(20, 6) \times 0.6\text{^}6 \times (1–0.6)\text{^}(20–6) \approx 0.0049$
$P(X = 7) = C(20, 7) \times 0.6\text{^}7 \times (1–0.6)\text{^}(20–7) \approx 0.0146$
$P(X < 8)$ is the sum of these individual values, which is 0.0210.

Therefore, the probability that fewer than 8 out of 20 randomly selected people prefer the particular brand of chocolate is approximately 0.0210 or 2.10%.

Example 10.48 A factory produces gears, and the probability of a gear being defective is 0.03. If the factory produces 100 gears, what is the probability that at least 5 gears are defective?

Given:

$n = 100$ (total number of trials)
$p = 0.03$ (probability of success, i.e., a defective gear)

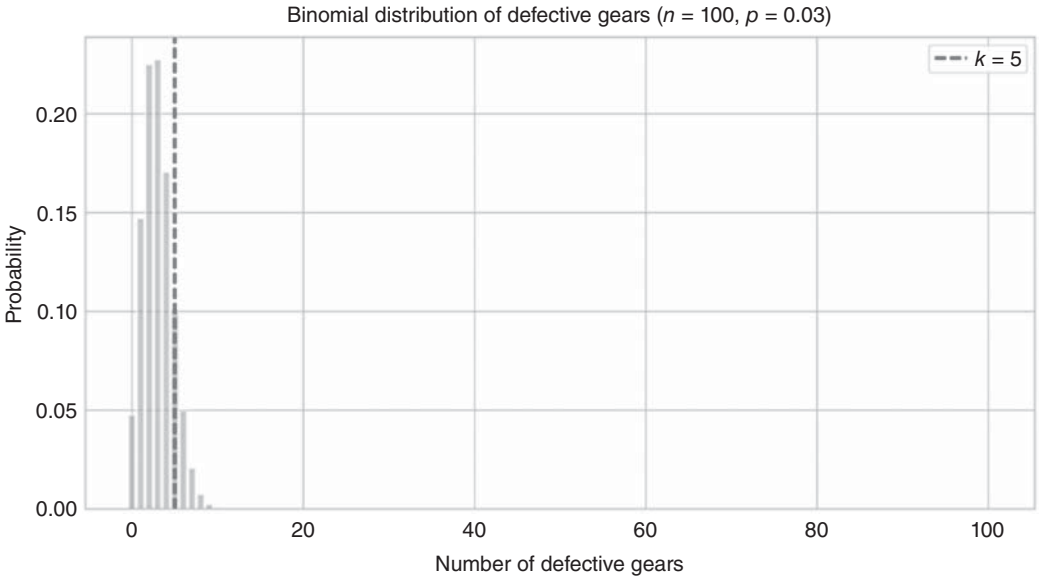
To find the probability that at least 5 gears are defective, we need to calculate the sum of probabilities for 5, 6, 7, ..., 99, and 100 defective gears:

$$P(X \geq 5) = 1 - P(X < 5) = 1 - (P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4))$$

Using the PMF formula for each term:

$$P(X = 0) = C(100, 0) \times 0.03\text{^}0 \times (1–0.03)\text{^}(100–0) \approx 0.0475$$
$$P(X = 1) = C(100, 1) \times 0.03\text{^}1 \times (1–0.03)\text{^}(100–1) \approx 0.1471$$
$$P(X = 2) = C(100, 2) \times 0.03\text{^}2 \times (1–0.03)\text{^}(100–2) \approx 0.2251$$
$$P(X = 3) = C(100, 3) \times 0.03\text{^}3 \times (1–0.03)\text{^}(100–3) \approx 0.2275$$
$$P(X = 4) = C(100, 4) \times 0.03\text{^}4 \times (1–0.03)\text{^}(100–4) \approx 0.1706$$
$$P(X \geq 5) = 1 - (0.0475 + 0.1471 + 0.2251 + 0.2275 + 0.1706) \approx 1–0.8178 = 0.1821$$

Therefore, the probability that at least 5 out of 100 gears produced by the factory are defective is approximately 0.1821 or 18.21%.



As was the case with the standard normal distribution, there exist tables of binomial probabilities, which can relieve us of the requirement to perform the computations shown above; naturally, there are also computer programs (such as Microsoft Excel) that can accelerate the process even further.

Let us re-examine two of our examples, this time using a binomial probability table.

Example 10.49 A fair coin is tossed 10 times. What is the probability of getting exactly six heads?
Solution:

$n = 10$ (total number of trials)
$k = 6$ (number of successes, i.e., heads)
$p = 0.5$ (probability of success, i.e., getting a head)

Solution:
Table 4 continued

		p																			
n	*r*	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
8	0	.923	.663	.430	.272	.168	.100	.058	.032	.017	.008	.004	.002	.001	.000	.000	.000	.000	.000	.000	.000
	1	.075	.279	.383	.385	.336	.267	.198	.137	.090	.055	.031	.016	.008	.003	.001	.000	.000	.000	.000	.000
	2	.003	.051	.149	.238	.294	.311	.296	.259	.209	.157	.109	.070	.041	.022	.010	.004	.001	.000	.000	.000
	3	.000	.005	.033	.084	.147	.208	.254	.279	.279	.257	.219	.172	.124	.081	.047	.023	.009	.003	.000	.000
	4	.000	.000	.005	.018	.046	.087	.136	.188	.232	.263	.273	.263	.232	.188	.136	.087	.046	.018	.005	.000
	5	.000	.000	.000	.003	.009	.023	.047	.081	.124	.172	.219	.257	.279	.279	.254	.208	.147	.084	.033	.005
	6	.000	.000	.000	.000	.001	.004	.010	.022	.041	.070	.109	.157	.209	.259	.296	.311	.294	.238	.149	.051
	7	.000	.000	.000	.000	.000	.000	.001	.003	.008	.016	.031	.055	.090	.137	.198	.267	.336	.385	.383	.279
	8	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.004	.008	.017	.032	.058	.100	.168	.272	.430	.663
9	0	.914	.630	.387	.232	.134	.075	.040	.021	.010	.005	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
	1	.083	.299	.387	.368	.302	.225	.156	.100	.060	.034	.018	.008	.004	.001	.000	.000	.000	.000	.000	.000
	2	.003	.063	.172	.260	.302	.300	.267	.216	.161	.111	.070	.041	.021	.010	.004	.001	.000	.000	.000	.000
	3	.000	.008	.045	.107	.176	.234	.267	.272	.251	.212	.164	.116	.074	.042	.021	.009	.003	.001	.000	.000
	4	.000	.001	.007	.028	.066	.117	.172	.219	.251	.260	.246	.213	.167	.118	.074	.039	.017	.005	.001	.000
	5	.000	.000	.001	.005	.017	.039	.074	.118	.167	.213	.246	.260	.251	.219	.172	.117	.066	.028	.007	.001
	6	.000	.000	.000	.001	.003	.009	.021	.042	.074	.116	.164	.212	.251	.272	.267	.234	.176	.107	.045	.008
	7	.000	.000	.000	.000	.000	.001	.004	.010	.021	.041	.070	.111	.161	.216	.267	.300	.302	.260	.172	.063
	8	.000	.000	.000	.000	.000	.000	.000	.001	.004	.008	.018	.034	.060	.100	.156	.225	.302	.368	.387	.299
	9	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.005	.010	.021	.040	.075	.134	.232	.387	.630
10	0	.904	.599	.349	.197	.107	.056	.028	.014	.006	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.091	.315	.387	.347	.268	.188	.121	.072	.040	.021	.010	.004	.002	.000	.000	.000	.000	.000	.000	.000
	2	.004	.075	.194	.276	.302	.282	.233	.176	.121	.076	.044	.023	.011	.004	.001	.000	.000	.000	.000	.000
	3	.000	.010	.057	.130	.201	.250	.267	.252	.215	.166	.117	.075	.042	.021	.009	.003	.001	.000	.000	.000
	4	.000	.001	.011	.040	.088	.146	.200	.238	.251	.238	.205	.160	.111	.069	.037	.016	.006	.001	.000	.000
	5	.000	.000	.001	.008	.026	.058	.103	.154	.201	.234	.246	.234	.201	.154	.103	.058	.026	.008	.001	.000
	6	.000	.000	.000	.001	.006	.016	.037	.069	.111	.160	.205	.238	.251	.238	.200	.146	.088	.040	.011	.001
	7	.000	.000	.000	.000	.001	.003	.009	.021	.042	.075	.117	.166	.215	.252	.267	.250	.201	.130	.057	.010
	8	.000	.000	.000	.000	.000	.000	.001	.004	.011	.023	.044	.076	.121	.176	.233	.282	.302	.276	.194	.07.
	9	.000	.000	.000	.000	.000	.000	.000	.000	.002	.004	.010	.021	.040	.072	.121	.188	.268	.347	.387	.315
	10	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.006	.014	.028	.056	.107	.197	.349	.599

The table entry corresponding to 6 successes in 10 trials, with a probability of success being 0.5, is 0.205; this can be computed using Microsoft Excel's BINOM.DIST command, with the field entries being the number of successes, number of trials, probability of success, and FALSE (false is the identifier used by Excel to indicate we are looking for a probability density value (rather than a cumulative probability density, which we will see in the next example).

The output of the Excel command BINOM.DIST(6,10,0.5,FALSE) is 0.205, as before.

Example 10.50 A basketball player has a free-throw success rate of 80%. If the player attempts 15 free throws, what is the probability of making at least 10 of them?

Given:

- $n = 15$ (total number of trials)
- $p = 0.8$ (probability of success, i.e., making a free throw)

To find the probability of making at least 10 free throws, we need to calculate the sum of probabilities for 10, 11, 12, 13, 14, and 15 successful attempts:

$$P(X \geq 10) = P(X = 10) + P(X = 11) + P(X = 12) + P(X = 13) + P(X = 14) + P(X = 15)$$

		p																		
n	r	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
11	6	.000	.000	.000	.002	.010	.027	.057	.099	.147	.193	.226	.236	.221	.183	.132	.080	.039	.013	.002	.000
	7	.000	.000	.000	.000	.002	.006	.017	.038	.070	.113	.161	.206	.236	.243	.220	.172	.111	.054	.016	.001
	8	.000	.000	.000	.000	.000	.001	.004	.010	.023	.046	.081	.126	.177	.225	.257	.258	.221	.152	.071	.014
	9	.000	.000	.000	.000	.000	.000	.001	.002	.005	.013	.027	.051	.089	.140	.200	.258	.295	.287	.213	.087
	10	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.005	.013	.027	.052	.093	.155	.236	.325	.384	.329
	11	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.004	.009	.020	.042	.086	.167	.314	.569
12	0	.886	.540	.282	.142	.069	.032	.014	.006	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.107	.341	.377	.301	.206	.127	.071	.037	.017	.008	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.006	.099	.230	.292	.283	.232	.168	.109	.064	.034	.016	.007	.002	.001	.000	.000	.000	.000	.000	.000
	3	.000	.017	.085	.172	.236	.258	.240	.195	.142	.092	.054	.028	.012	.005	.001	.000	.000	.000	.000	.000
	4	.000	.002	.021	.068	.133	.194	.231	.237	.213	.170	.121	.076	.042	.020	.008	.002	.001	.000	.000	.000
	5	.000	.000	.004	.019	.053	.103	.158	.204	.227	.223	.193	.149	.101	.059	.029	.011	.003	.001	.000	.000
	6	.000	.000	.000	.004	.016	.040	.079	.128	.177	.212	.226	.212	.177	.128	.079	.040	.016	.004	.000	.000
	7	.000	.000	.000	.001	.003	.011	.029	.059	.101	.149	.193	.223	.227	.204	.158	.103	.053	.019	.004	.000
	8	.000	.000	.000	.000	.001	.002	.008	.020	.042	.076	.121	.170	.213	.237	.231	.194	.133	.068	.021	.002
	9	.000	.000	.000	.000	.000	.000	.001	.005	.012	.028	.054	.092	.142	.195	.240	.258	.236	.172	.085	.017
	10	.000	.000	.000	.000	.000	.000	.000	.001	.002	.007	.016	.034	.064	.109	.168	.232	.283	.292	.230	.099
	11	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.008	.017	.037	.071	.127	.206	.301	.377	.341
	12	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.002	.006	.014	.032	.069	.142	.282	.540
15	0	.860	.463	.206	.087	.035	.013	.005	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1	.130	.366	.343	.231	.132	.067	.031	.013	.005	.002	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	2	.009	.135	.267	.286	.231	.156	.092	.048	.022	.009	.003	.001	.000	.000	.000	.000	.000	.000	.000	.000
	3	.000	.031	.129	.218	.250	.225	.170	.111	.063	.032	.014	.005	.002	.000	.000	.000	.000	.000	.000	.000
	4	.000	.005	.043	.116	.188	.225	.219	.179	.127	.078	.042	.019	.007	.002	.001	.000	.000	.000	.000	.000
	5	.000	.001	.010	.045	.103	.165	.206	.212	.186	.140	.092	.051	.024	.010	.003	.001	.000	.000	.000	.000
	6	.000	.000	.002	.013	.043	.092	.147	.191	.207	.191	.153	.105	.061	.030	.012	.003	.001	.000	.000	.000
	7	.000	.000	.000	.003	.014	.039	.081	.132	.177	.201	.196	.165	.118	.071	.035	.013	.003	.001	.000	.000
	8	.000	.000	.000	.001	.003	.013	.035	.071	.118	.165	.196	.201	.177	.132	.081	.039	.014	.003	.000	.000
	9	.000	.000	.000	.000	.001	.003	.012	.030	.061	.105	.153	.191	.207	.191	.147	.092	.043	.013	.002	.000
	10	.000	.000	.000	.000	.000	.001	.003	.010	.024	.051	.092	.140	.186	.212	.206	.165	.103	.045	.010	.001
	11	.000	.000	.000	.000	.000	.000	.001	.002	.007	.019	.042	.078	.127	.179	.219	.225	.188	.116	.043	.005
	12	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.014	.032	.063	.111	.170	.225	.250	.218	.129	.031
	13	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.003	.009	.022	.048	.092	.156	.231	.286	.267	.135
	14	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.013	.031	.067	.132	.231	.343	.366
	15	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.005	.013	.035	.087	.206	.463

We see that, for $n = 20$ and $p = 0.80$, the sum of the table entries corresponding to 10, 11, 12, 13, 14, and 15 successes is: $0.103 + 0.188 + 0.250 + 0.231 + 0.132 + 0.035 = 0.939$.

Using Microsoft Excel's **BINOM.DIST.RANGE** function, whose field entries are number of trials, probability of success, lower range of successes, upper range of successes, we find:

BINOM.DIST.RANGE(15,0.8,10,15) ≈ 0.939, which agrees with our computational results.

10.7 Linear Correlation and Regression

Linear correlation and regression are statistical techniques used to study the possibility of a linear relationship between two continuous variables. Correlation measures the strength and direction of the linear relationship between two variables, while regression allows us to construct an equation with which we can predict the value of one variable based on the value of the other. In this section, we will discuss the concepts of correlation coefficient, least squares regression line, and their applications.

10.7.1 Correlation Coefficient

The correlation coefficient, denoted by r, is a measure of the strength and direction of the linear relationship between two variables, X and Y. The value of r ranges from -1 to $+1$, where:

- $r = 1$ indicates a perfect positive linear relationship
- $r = -1$ indicates a perfect negative linear relationship
- $r = 0$ indicates no linear relationship

The formula for the correlation coefficient is:

$$r = (\Sigma((X - \overline{X})(Y - \overline{Y})))/(\sqrt{(\Sigma(X - \overline{X})^2)} \times \sqrt{(\Sigma(Y - \overline{Y})^2)})$$

where

- $\overline{X}$ is the mean of X values
- $\overline{Y}$ is the mean of Y values

The values of "r" lie on a relative scale. The closer $|r|$ is to 1, the stronger the correlation. In physics and chemistry, $|r|$ should be higher than 0.9 for the correlation to be considered meaningful, while in social sciences, the threshold for significance could be as low as 0.5.

Example 10.51 Calculate and interpret the value of the correlation coefficient between the number of hours studied and the test scores for a group of students.

Hours Studied (X): 2, 3, 5, 6, 8
Test Scores (Y): 60, 70, 80, 90, 95

Step 1: Calculate the means of X and Y.

$\overline{X} = (2 + 3 + 5 + 6 + 8) / 5 = 4.8$
$\overline{Y} = (60 + 70 + 80 + 90 + 95) / 5 = 79$

Step 2: Calculate $(X - \overline{X})$, $(Y - \overline{Y})$, $(X - \overline{X})^2$, and $(X - \overline{X})(Y - \overline{Y})$.

$(X - \overline{X})$: $-2.8, -1.8, 0.2, 1.2, 3.2$
$(Y - \overline{Y})$: $-19, -9, 1, 11, 16$
$(X - \overline{X})^2$: 7.84, 3.24, 0.04, 1.44, 10.24
$(X - \overline{X})(Y - \overline{Y})$: 53.2, 16.2, 0.2, 13.2, 51.2

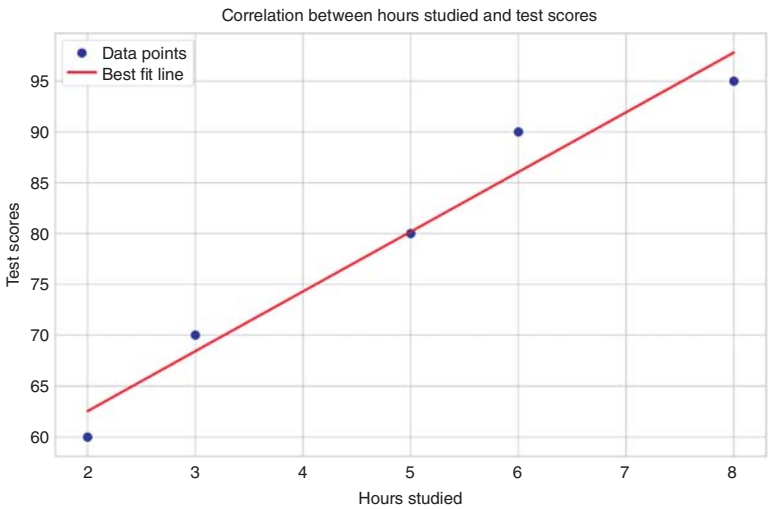
Step 3: Calculate $\Sigma((X - \overline{X})(Y - \overline{Y}))$, $\Sigma(X - \overline{X})^2$, and $\Sigma(Y - \overline{Y})^2$.

$\Sigma((X - \overline{X})(Y - \overline{Y})) = 53.2 + 16.2 + 0.2 + 13.2 + 51.2 = 134$
$\Sigma(X - \overline{X})^2 = 7.84 + 3.24 + 0.04 + 1.44 + 10.24 = 22.8$
$\Sigma(Y - \overline{Y})^2 = 361 + 81 + 1 + 121 + 256 = 820$

Step 4: Calculate the correlation coefficient.

$$r = 134 / (\sqrt{(22.8)} \times \sqrt{(820)}) \approx 0.99$$

The correlation coefficient of 0.99 indicates a strong positive linear relationship between the number of hours studied and test scores.



10.7.2 Least Squares Regression Line

The least squares regression line is the line that "best fits" the data points in a scatter plot. That is, if the data points arrange themselves nearly or exactly upon a straight-line path in the plane, it is reasonable to find a line that most nearly passes through the center of the data points. We call this line the "line of best fit."

The line of best fit minimizes the sum of the squared vertical distances between the data points and the line. Its equation, using the slope–intercept form of the equation of a line, is:

$$\hat{y} = b_0 + b_1 x$$

where

- $\hat{y}$ is the predicted value of Y for a given value of X
- b_0 is the y-intercept of the regression line
- b_1 is the slope of the regression line

The formulas for b_0 and b_1 are:

$$b_1 = (\Sigma((X - \overline{X})(Y - \overline{Y}))) / (\Sigma(X - \overline{X})^2)$$
$$b_0 = \overline{Y} - b1\text{^}\overline{X}$$

It is important to realize the use of this linear equation as a reliable predictor is restricted to the range of the data. Moving past that range takes us to the realm of "extrapolation," which (though it can be useful) is not recommended.

Example 10.52 Find the equation of the least squares regression line for the data in Example 1.

Using the values calculated in Example 10.51:

$$b_1 = 134 / 22.8 \approx 5.88$$
$$b_0 = 79 - (5.88 \times 4.8) \approx 50.78$$

The equation of the least squares regression line is:

$$\hat{y} = 50.78 + 5.88x$$

This equation can be used to predict the test score (Y) for a given number of hours studied (X), provided the hours of study lie between 2 and 8. Outside that span of hours, the equation is less reliable as a predictor.

For example, imagine an individual studies for 3.75 hours for the exam; what score can be expected, according to the linear regression equation?

$$\hat{y} = 50.78 + 5.88(3.75) = 50.78 + 22.05 = 72.83.$$

Now, imagine an individual studies for 14 hours. This is outside the span of values for which the equation can be expected to be reliable, but what would it tell us?

$$\begin{aligned}\hat{y} &= 50.78 + 5.88(14) \\ &= 50.78 + 82.32 \\ &= 133.10\end{aligned}$$

As you can see, this exceeds the total possible for the examination. Our ability to forecast beyond the range of values given in the dataset is unreliable (this is an illustration of this shortcoming).

Example 10.53 Calculate the correlation coefficient between age and systolic blood pressure of a group of individuals.

Age (X): 25, 30, 35, 40, 45, 50, 55
Systolic Blood Pressure (Y): 120, 125, 130, 135, 140, 145, 150

Step 1: Calculate the means of X and Y.

$$\overline{X} = (25 + 30 + 35 + 40 + 45 + 50 + 55) / 7 = 40$$
$$\overline{Y} = (120 + 125 + 130 + 135 + 140 + 145 + 150) / 7 = 135$$

Step 2: Calculate $(X - \overline{X})$, $(Y - \overline{Y})$, $(X - \overline{X})^2$, and $(X - \overline{X})(Y - \overline{Y})$.

$(X - \overline{X})$: $-15, -10, -5, 0, 5, 10, 15$
$(Y - \overline{Y})$: $-15, -10, -5, 0, 5, 10, 15$
$(X - \overline{X})^2$: 225, 100, 25, 0, 25, 100, 225
$(X - \overline{X})(Y - \overline{Y})$: 225, 100, 25, 0, 25, 100, 225

Step 3: Calculate $\Sigma((X - \overline{X})(Y - \overline{Y}))$, $\Sigma(X - \overline{X})^2$, and $\Sigma(Y - \overline{Y})^2$.

$$\Sigma((X - \overline{X})(Y - \overline{Y})) = 225 + 100 + 25 + 0 + 25 + 100 + 225 = 700$$
$$\Sigma(X - \overline{X})^2 = 225 + 100 + 25 + 0 + 25 + 100 + 225 = 700$$
$$\Sigma(Y - \overline{Y})^2 = 225 + 100 + 25 + 0 + 25 + 100 + 225 = 700$$

Step 4: Calculate the correlation coefficient.

$$r = 700 / (\sqrt{(700)} \times \sqrt{(700)}) = 1$$

The correlation coefficient of 1 indicates a perfect positive linear relationship between age and systolic blood pressure.

Example 10.54 Find the equation of the least squares regression line for the data in the blood pressure example.

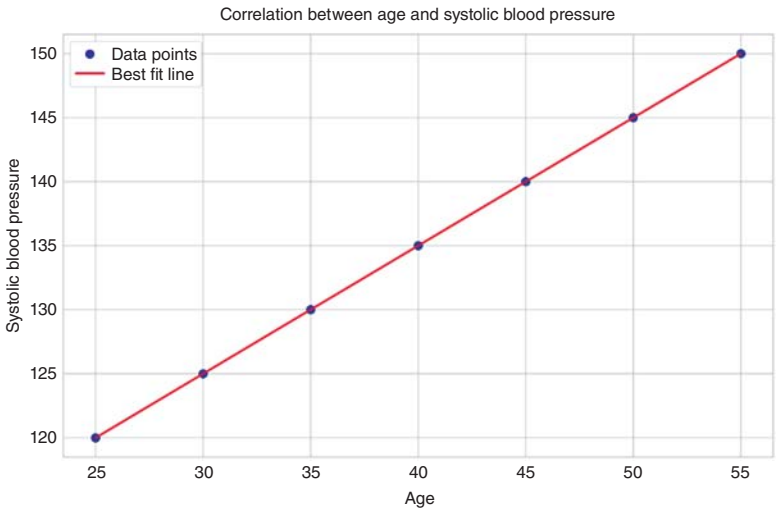
Using the values calculated in Example 3:

$$b_1 = 700 / 700 = 1$$
$$b_0 = 135 - (1 \times 40) = 95$$

The equation of the least squares regression line is:

$$\hat{y} = 95 + x$$

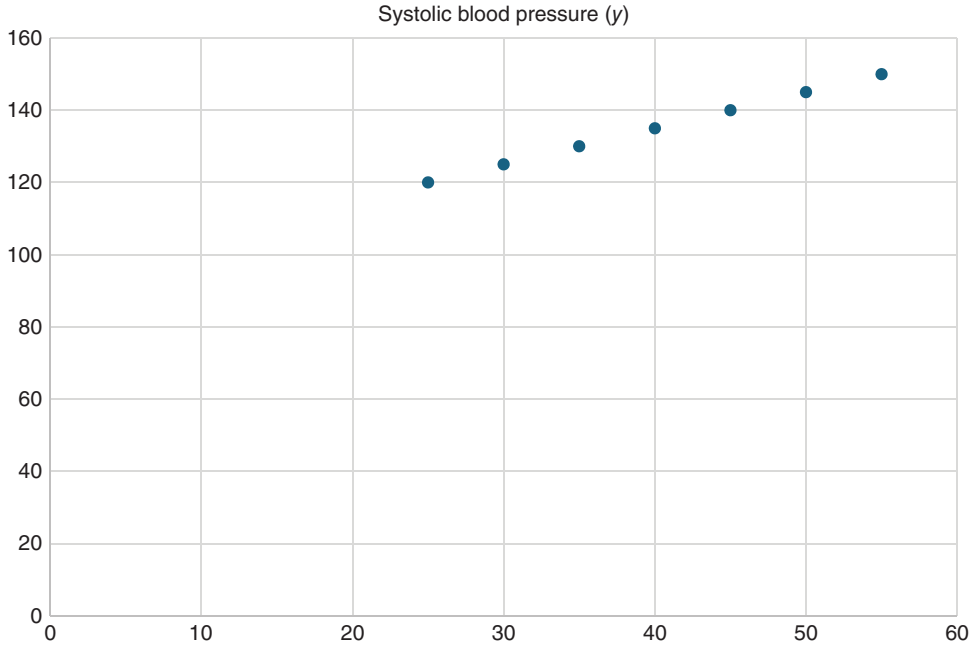
This equation can be used to predict the systolic blood pressure (Y) for a given age (X).



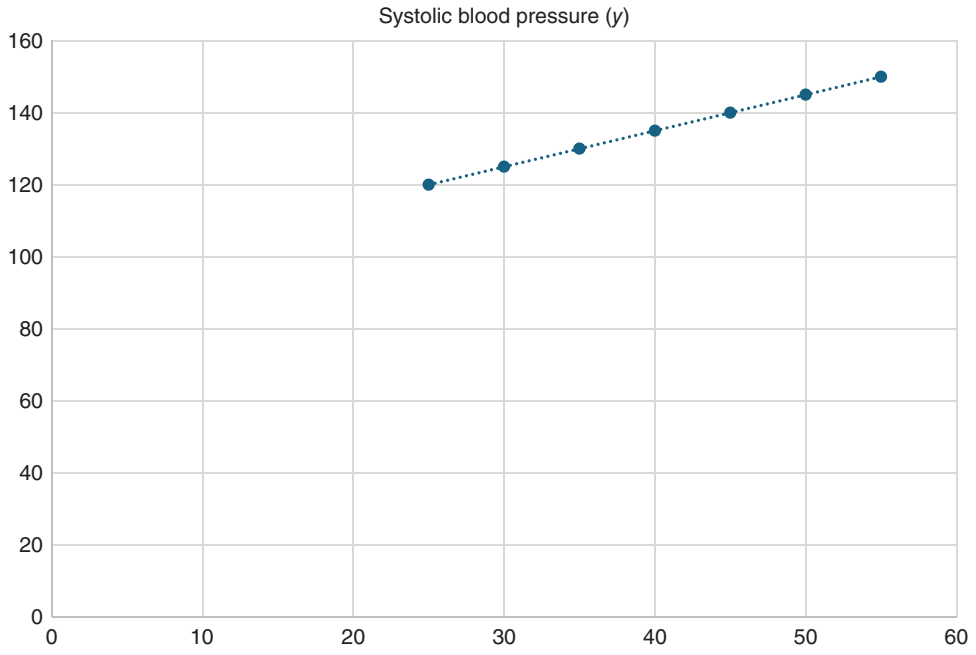
As we saw in the normal and binomial distributions, technology is useful as an aid to quicken our progression through the straightforward but tedious computations we have seen above.

Let us consider the example just shown (relating age and blood pressure) and use the TREND-LINE feature of Excel to produce our result with greater ease.

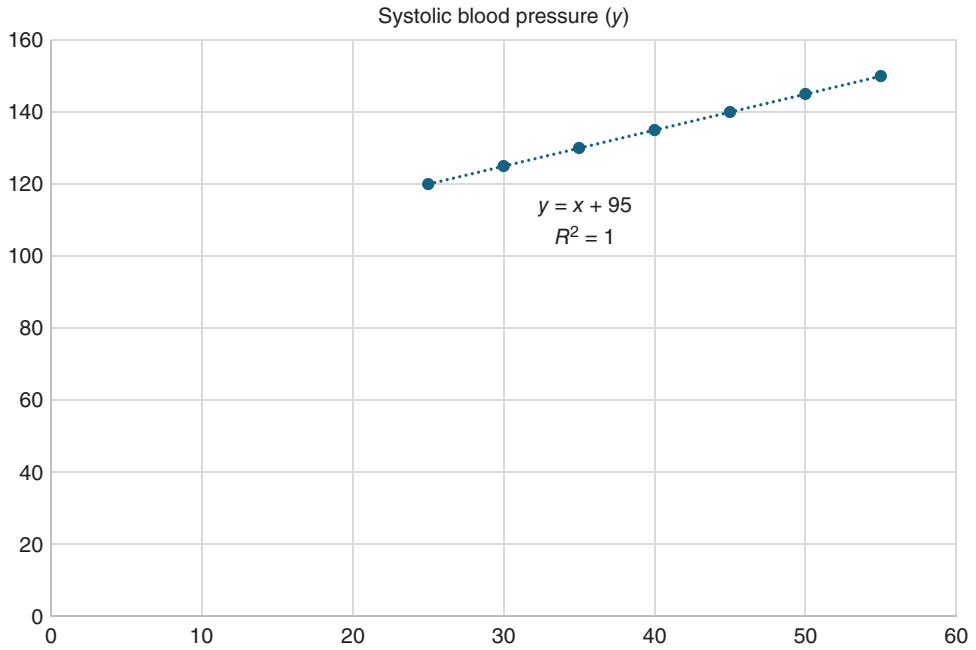
First, we create a spreadsheet and enter our data. Then, we create a scatter plot of that data:



If we mouse-click on one of the data points plotted, all the data is automatically selected, and we can then right-click and choose "Add Trendline."



Using the formatting tools of Excel, we can insert the equation of the line and the value of r^2, which is:



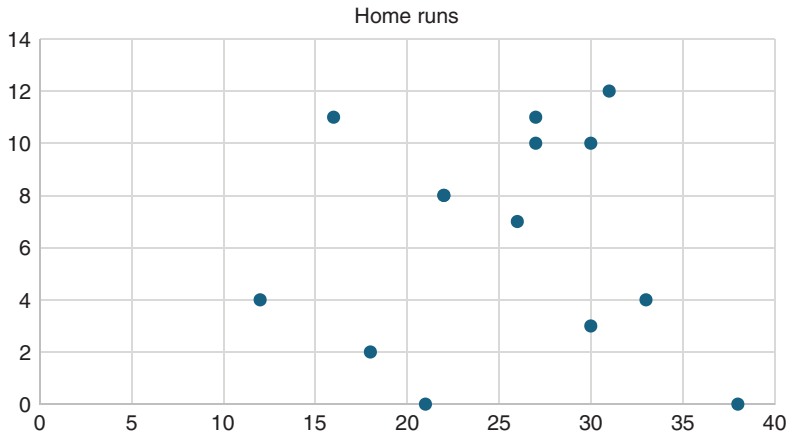
Note the equation is the same as we found earlier, and we encounter the extremely unusual (and thus, likely contrived) result that there is perfect positive linear correlation between age and systolic blood pressure.

When correlation is perfect, our suspicions should be elevated. Unless the data is constructed (such as in an example like this) in such a way as to deliberately produce perfect positive linear correlation, we would be astounded to find such to be the case.

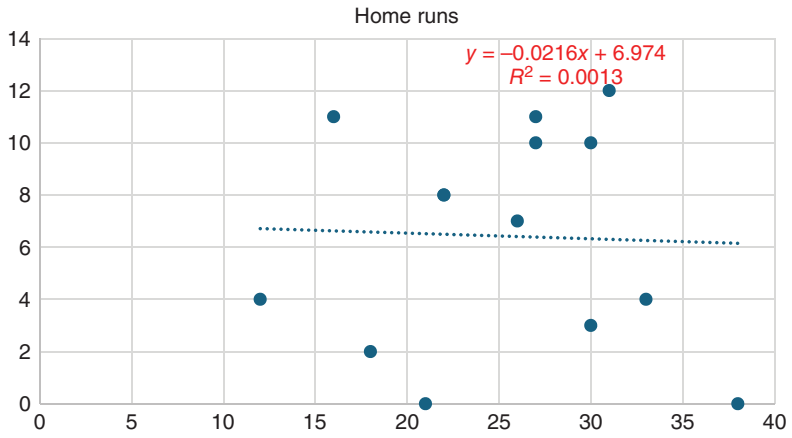
Example 10.55 A high school baseball coach believes there is a correlation between the number of pushups a player can perform in a minute and the number of home runs they will hit in a season. He compiles data for 14 players on the varsity team and finds:

Pushups	Home Runs
22	8
18	2
21	0
16	11
27	10
33	4
30	3
12	4
38	0
31	12
27	11
22	8
26	7
30	10

The data is graphed, and it is seen:



There seems little evidence a linear correlation exists between pushups and home runs (the data does not appear to follow a linear pattern), but we use Excel to find the best-fit line and the correlation coefficient:



We see that the correlation is negative (the trendline has a negative slope) and that $r \approx -0.036$. There is an extremely weak negative linear correlation present.

If we use the equation of the best-fit line, regardless of the weakness of the relationship, and wonder how many home runs a player can be expected to hit if they can perform 29 pushups, we find:

$$\hat{y} = -0.0216(29) + 6.974 \approx 6.35$$

The player is expected to hit approximately six home runs during the season.

Example 10.56 A company wants to analyze the relationship between its advertising expenditure and sales. The data collected over the past six months is as follows:

Advertising Expenditure (X) (in thousands of dollars): 10, 15, 20, 25, 30, 35

Sales (Y) (in thousands of units): 50, 60, 70, 80, 90, 100

Calculate the correlation coefficient and find the equation of the least squares regression line.
Step 1: Calculate the means of X and Y.

$$\overline{X} = (10 + 15 + 20 + 25 + 30 + 35) / 6 = 22.5$$
$$\overline{Y} = (50 + 60 + 70 + 80 + 90 + 100) / 6 = 75$$

Step 2: Calculate $(X - \overline{X})$, $(Y - \overline{Y})$, $(X - \overline{X})^2$, and $(X - \overline{X})(Y - \overline{Y})$.

$(X - \overline{X})$: $-12.5, -7.5, -2.5, 2.5, 7.5, 12.5$
$(Y - \overline{Y})$: $-25, -15, -5, 5, 15, 25$
$(X - \overline{X})^2$: 156.25, 56.25, 6.25, 6.25, 56.25, 156.25
$(X - \overline{X})(Y - \overline{Y})$: 312.5, 112.5, 12.5, 12.5, 112.5, 312.5

Step 3: Calculate $\Sigma((X - \overline{X})(Y - \overline{Y}))$, $\Sigma(X - \overline{X})^2$, and $\Sigma(Y - \overline{Y})^2$.

$$\Sigma((X - \overline{X})(Y - \overline{Y})) = 312.5 + 112.5 + 12.5 + 12.5 + 112.5 + 312.5 = 875$$
$$\Sigma(X - \overline{X})^2 = 156.25 + 56.25 + 6.25 + 6.25 + 56.25 + 156.25 = 437.5$$
$$\Sigma(Y - \overline{Y})^2 = 625 + 225 + 25 + 25 + 225 + 625 = 1750$$

Step 4: Calculate the correlation coefficient.

$$r = 875 / (\sqrt{(437.5)} \times \sqrt{(1750)}) \approx 0.9999$$

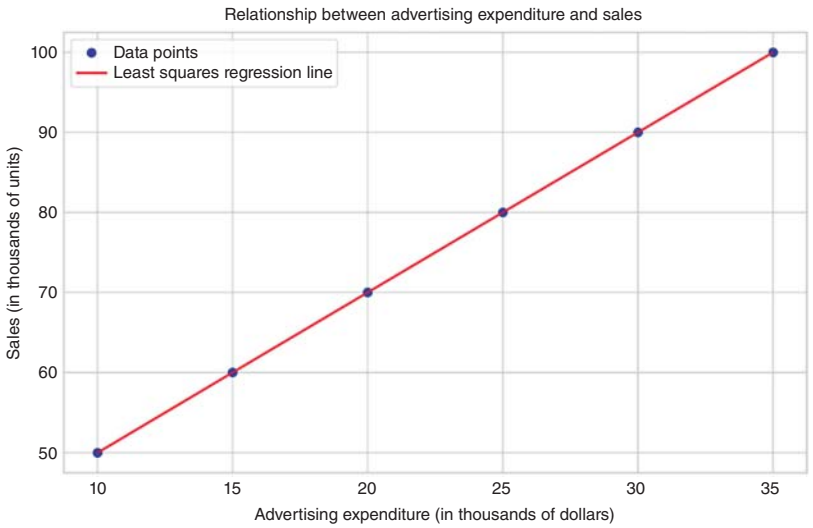
The correlation coefficient of 0.9999 indicates a very strong positive linear relationship between advertising expenditure and sales.
Step 5: Find the equation of the least squares regression line.

$$b_1 = 875 / 437.5 = 2$$
$$b_0 = 75 - (2 \times 22.5) = 30$$

The equation of the least squares regression line is:

$$\hat{y} = 30 + 2x$$

This equation can be used to predict the sales (Y) for a given advertising expenditure (X).



Example 10.57 A study was conducted to investigate the relationship between the number of hours of sleep per night and the level of productivity the following day. The data collected from eight participants is as follows:

Hours of Sleep (X): 5, 6, 6, 7, 7, 8, 8, 9
Productivity Score (Y): 60, 65, 70, 75, 80, 85, 90, 95
Calculate the correlation coefficient and find the equation of the least squares regression line.

Step 1: Calculate the means of X and Y.

$$\overline{X} = (5+6+6+7+7+8+8+9) / 8 = 7$$
$$\overline{Y} = (60+65+70+75+80+85+90+95) / 8 = 77.5$$

Step 2: Calculate $(X-\overline{X})$, $(Y-\overline{Y})$, $(X-\overline{X})^2$, and $(X-\overline{X})(Y-\overline{Y})$.

$(X-\overline{X})$: $-2, -1, -1, 0, 0, 1, 1, 2$
$(Y-\overline{Y})$: $-17.5, -12.5, -7.5, -2.5, 2.5, 7.5, 12.5, 17.5$
$(X-\overline{X})^2$: 4, 1, 1, 0, 0, 1, 1, 4
$(X-\overline{X})(Y-\overline{Y})$: 35, 12.5, 7.5, 0, 0, 7.5, 12.5, 35

Step 3: Calculate $\Sigma((X-\overline{X})(Y-\overline{Y}))$, $\Sigma(X-\overline{X})^2$, and $\Sigma(Y-\overline{Y})^2$.

$$\Sigma((X-\overline{X})(Y-\overline{Y})) = 35+12.5+7.5+0+0+7.5+12.5+35 = 110$$
$$\Sigma(\mathrm{X}-\overline{X})^2 = 4+1+1+0+0+1+1+4 = 12$$
$$\Sigma(Y-\overline{Y})^2 = 306.25+156.25+56.25+6.25+6.25+56.25+156.25+306.25 = 1050$$

Step 4: Calculate the correlation coefficient.

$$r = 110 / (\sqrt{(12)} \times \sqrt{(1050)}) \approx 0.9805$$

The correlation coefficient of 0.9805 indicates a strong positive linear relationship between hours of sleep and productivity.

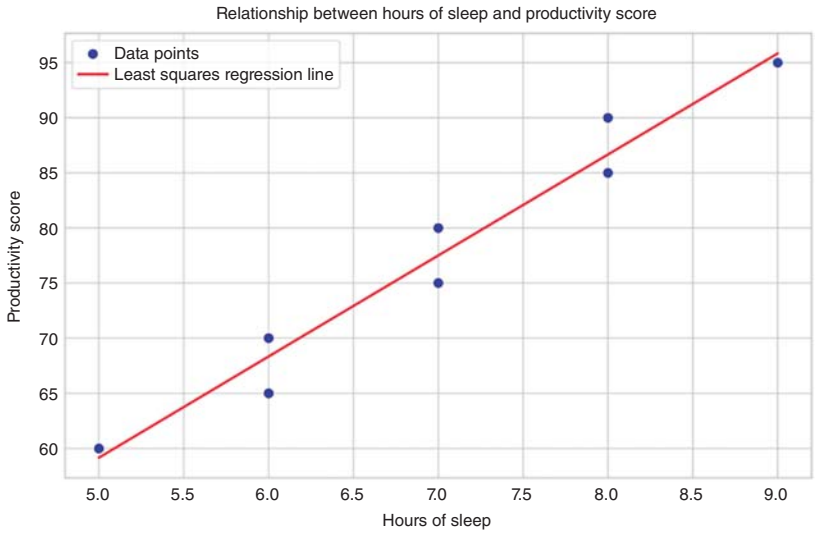
Step 5: Find the equation of the least squares regression line.

$$b_1 = 110 / 12 \approx 9.17$$
$$b_0 = 77.5 - (9.17 \times 7) = 13.31$$

The equation of the least squares regression line is:

$$\hat{y} = 13.31 + 9.17x$$

This equation can be used to predict the productivity score (Y) for a given number of hours of sleep (X).



Example 10.58 A fitness tracker company wants to analyze the relationship between the number of daily steps taken and the number of calories burned. The data collected from 10 users is as follows:

Daily Steps (X) (in thousands): 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Calories Burned (Y): 200, 250, 300, 350, 400, 450, 500, 550, 600, 650
Calculate the correlation coefficient and find the equation of the least squares regression line.

Step 1: Calculate the means of X and Y.

$\overline{X} = (2+3+4+5+6+7+8+9+10+11) / 10 = 6.5$
$\overline{Y} = (200+250+300+350+400+450+500+550+600+650) / 10 = 425$

Step 2: Calculate $(X-\overline{X})$, $(Y-\overline{Y})$, $(X-\overline{X})^2$, and $(X-\overline{X})(Y-\overline{Y})$.

$(X-\overline{X})$: $-4.5, -3.5, -2.5, -1.5, -0.5, 0.5, 1.5, 2.5, 3.5, 4.5$
$(Y-\overline{Y})$: $-225, -175, -125, -75, -25, 25, 75, 125, 175, 225$
$(X-\overline{X})^2$: 20.25, 12.25, 6.25, 2.25, 0.25, 0.25, 2.25, 6.25, 12.25, 20.25
$(X-\overline{X})(Y-\overline{Y})$: 1012.5, 612.5, 312.5, 112.5, 12.5, 12.5, 112.5, 312.5, 612.5, 1012.5

Step 3: Calculate $\Sigma((X-\overline{X})(Y-\overline{Y}))$, $\Sigma(X-\overline{X})^2$, and $\Sigma(Y-\overline{Y})^2$.

- $\Sigma((X-\overline{X})(Y-\overline{Y})) = 1012.5+612.5+312.5+112.5+12.5+12.5+112.5+312.5+612.5+1012.5 = 4125$
- $\Sigma(X-\overline{X})^2 = 20.25+12.25+6.25+2.25+0.25+0.25+2.25+6.25+12.25+20.25 = 82.5$
- $\Sigma(Y-\overline{Y})^2 = 50\,625+30\,625+15\,625+5625+625+625+5625+15\,625+30\,625+50\,625 = 206\,250$

Step 4: Calculate the correlation coefficient.

- $r = 4125 / (\sqrt{(82.5)} \times \sqrt{(206\,250)}) = 1$

The correlation coefficient of 1 indicates a perfect positive linear relationship between daily steps and calories burned.

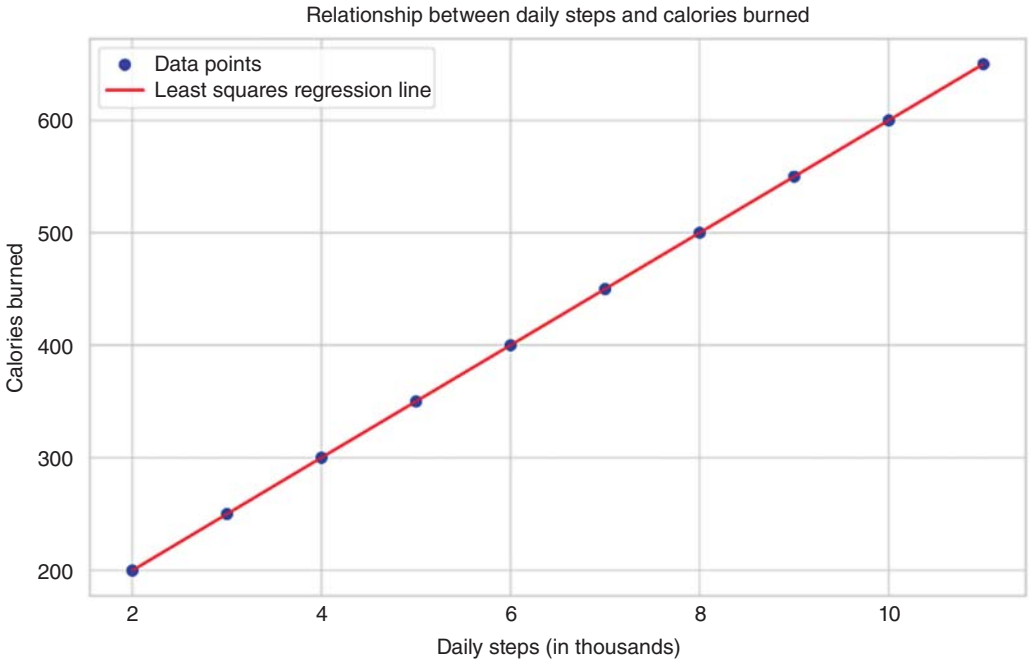
Step 5: Find the equation of the least squares regression line.

- $b_1 = 4125 / 82.5 = 50$
- $b_0 = 425 - (50 \times 6.5) = 100$

The equation of the least squares regression line is:

- $\hat{y} = 100 + 50x$

This equation can be used to predict the number of calories burned (Y) for a given number of daily steps (X).


Relationship between daily steps and calories burned
Data points
Least squares regression line
Calories burned
600
500
400
300
200
2
4
6
8
10
Daily steps (in thousands)

11

Basic Concepts of Probability

Imagine you are tossing a fair coin. What is the probability that it lands heads up? You might think the answer is obvious—50%. But probability is more than just guessing outcomes. It's a way of understanding the likelihood of various outcomes for events that involve uncertainty (the existence of several different results), and plays a crucial role in decision-making, statistics, science, and everyday life. Whether you are trying to predict the weather, assessing the odds of winning a game, or calculating risks in business, probability is the language of uncertainty.

In a sense, probability is a *predictive science*. By this, we mean we are looking ahead in time and asking, "How likely is a particular outcome to occur?"

In this chapter, we will explore the foundational concepts of probability, starting with simple experiments like coin flips and dice rolls. We will learn to calculate the likelihood of events happening, the difference between independent and dependent events, and how probability helps us make sense of the world in situations where randomness is involved.

By the end of this chapter, you'll be equipped with the basic tools to begin thinking probabilistically—an essential skill for anyone working in fields such as finance, engineering, computer science, or even genetics.

11.1 Experiments, Outcomes, Sample Spaces, and Events

Probability is a mathematical tool used to study randomness. It deals with the chance (or likelihood) of one of several potential outcomes or events occurring in a given circumstance. At its core are the fundamental concepts of experiments, sample spaces, events, and successful versus unsuccessful outcomes. Probability offers a numerical way of describing how likely an event is to occur (or, conversely, not to occur) and is a measure associated with how certain we are of an outcome of an experiment or activity.

An **experiment** is an investigation carried out under controlled conditions. If the result is not predetermined, then the outcome is an illustration of **chance**. Rolling a pair of 12-sided dice is an example of an experiment.

> Note: You are likely to be familiar with six-sided dice. There are, particularly in tabletop role-playing games, dice having different numbers of sides, such as 4, 8, 10, 12, and 20. Such dice are referred to as polyhedral dice.

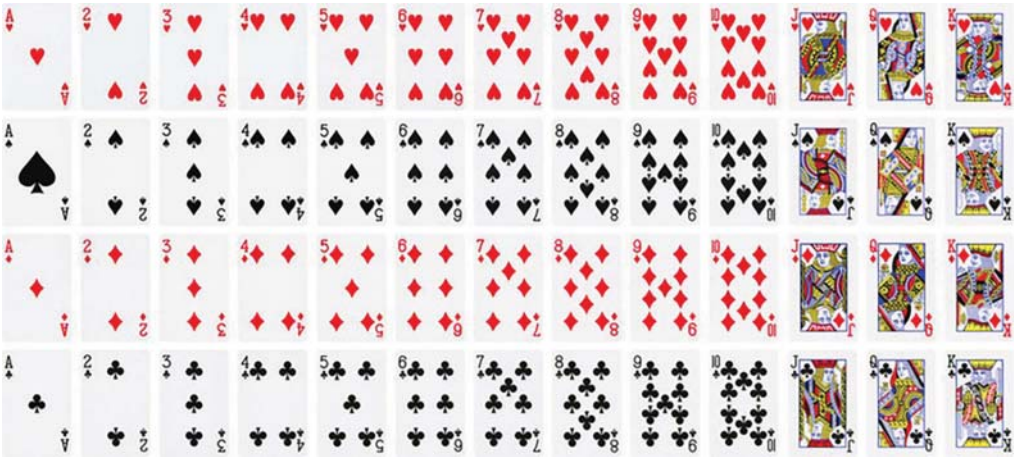
Another example of an experiment for which the outcome is a matter of chance is the selection of a single card from a 52-card, well-shuffled deck of playing cards. Assuming the cards have been "randomized" through the shuffling, drawing a card from the deck is an action having 52 distinct possible outcomes. We mention here that the instantiations of experiments are considered "fair."

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

In other words, there is no effort to manipulate the results; examples of this would be "loaded dice," or "stacked decks of cards," which would influence the likelihood of particular outcomes. Conditions that could initiate results that are not "fair" could be accidental; dice created in a factory using a plastic injection process could be "unbalanced" in such a manner that faces of the two dice are not equally-likely to occur.

If an experiment is to be performed multiple times, we refer to each attempt as a **trial**. The result of each trial is said to be an **outcome** of the experiment, and the set of all possible outcomes of an experiment is called the **sample space**.

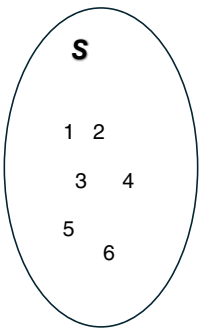
The sample space for rolling a single six-sided die would be the set $S = \{1, 2, 3, 4, 5, 6\}$. For flipping a coin, the sample space is $S = \{H, T\}$, where H represents "heads," and T represents "tails". For drawing a card from a standard deck of 52 cards, the sample space is the set of all 52 cards.



Source: MMphotos / Adobe Stock Photos.

Listing the possible outcomes is one method of visualizing the sample space. If the sample space is relatively small (admittedly, this is a subjective term), this can be an effective method for illustrating the possible outcomes of a trial of an experiment.

Small sample spaces can also be depicted using a diagram. For the example of rolling a six-sided die, this would look like:



There are times when visual aids such as a graph of this type can be useful, but one can easily see that (for a large sample space) this visual approach would become unwieldy. Imagine we were drawing numbers from a set ranging from 0 to 999; a visual representation of the set S would be impractical due to the sheer size of the set.

An **event** is a collection of one or more specific outcomes that share a particular characteristic. It is essentially a subset of the possible outcomes within an experiment, meaning that an event is made up of one or more outcomes.

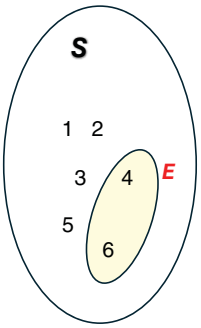
Example 11.1 Provide an example of an event E for each of the following scenarios (the answer is not unique):

- Rolling an eight-sided die
- Flipping a coin
- Drawing a card from a deck of 52 randomized cards

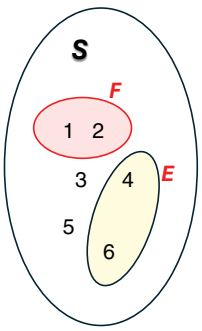
Solution:

1) When rolling an eight-sided die, an illustration of an event is rolling an even number on the die; we can represent this with the set $E = \{2, 4, 6, 8\}$.
2) When flipping a coin, the event of getting heads would be the set $E = \{H\}$.
3) When drawing a card from a deck, the event of drawing a face card would be the set $E = \{JH, JS, JC, JD, QH, QS, QC, QD, KH, KS, KC, KD\}$, where J represents a jack, Q represents a queen, K represents a king, H indicates hearts, S indicates spades, C indicates clubs, and D represents diamonds.

Returning to our graph of the sample space showing the outcomes for rolling a six-sided die, the event $E = \{4, 6\}$ can be shown as follows:



In probability theory, two events are said to be **mutually exclusive** (or disjoint) if they share no outcomes. For example, consider a new event F as shown below. The events share no outcomes from the sample space; thus, E and F are mutually exclusive events:



An event having a single possible outcome is referred to as a **simple event.** If we consider our previous example of tossing a coin, we get one outcome that is a head or a tail. A **compound event**

is more complex than a simple event, as it involves more than one outcome. Another way to view compound events is as *a combination of two or more simple events*.

Consider the event of rolling a 10-sided die and obtaining an even number less than 5. The outcomes being described are 2 and 4, so the event is $E = \{2, 4\}$. It is worthwhile to pause a moment to consider the clarity of language. Be mindful of how small distinctions can be significant! For instance, imagine one set of real numbers is the positive reals, while a second set is the negative reals. At first glance we might think this "covers" the set of all real numbers. However, reflection exposes the fact that 0 is neither positive nor negative; thus, 0 is not included in either of the sets described a moment ago.

The **complement of an event E**, which we will denote as E^c, is that part of the sample space outside event E. For the 10-sided die example, $E^c = \{1, 3, 5, 6, 7, 8, 9, 10\}$. The union of E and E^c is equal to the entire sample space, and the complement of the sample space is the empty set.

By construction, E and E^c are mutually exclusive.

Example 11.2 With a 52-card deck of cards, let E represent the event of drawing an ace from the deck. What is the complement of event E?

Solution:
One way of expressing E^c is the act of drawing a card from the deck that is not an ace. Another way we could identify E^c is the act of drawing a 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K from the deck. The verbal descriptions may differ, but they express precisely the same notion.

Success and Failure
In a probability experiment, a particular outcome or set of outcomes will be identified as a **success** (favorable, or desirable). Results not conforming to the identified notion of success are referred to as a **failure**.

Example 11.3 Imagine we have an opaque box in which there are 13 black marbles, 10 white marbles, 11 yellow marbles, and 6 blue marbles. Further, suppose we are interested in drawing a blue marble from the box; what is considered a success, and what is considered a failure?

Solution:
The phrasing of the problem indicates the act of drawing a blue marble from the box is considered a "success." The act of drawing a non-blue marble from the box is considered a failure.

Example 11.4 In a poll of 75 residents of the small island nation of Togoland, those individuals who had a favorable view of charismatic leader Maraca Kabob numbered 31, while those who had an unfavorable view numbered 44. If the pollsters are attempting to determine the likelihood a given resident had an unfavorable view of the island leader, what result is considered a success in this problem?

Solution:
Because we are interested in the feature of NOT having a positive view of Maraca Kabob, a success here would be finding a person who does not have a favorable opinion of the island leader.

11.1.1 Theoretical Probability

Theoretical probability is the likelihood of an event occurring based on known or assumed facts, often calculated using mathematical principles. It is determined by dividing the number of

favorable outcomes (successes) by the total number of possible outcomes in a given sample space. This type of probability is often used when the experiment or situation is idealized or perfectly controlled, and assuming all outcomes are equally likely.

The formula defining the **theoretical probability** of an event E is calculated as:

$$\mathrm{P}(E) = |E|/|S|$$

where $|E|$ is the number of elements in the event E, and $|S|$ is the number of elements in the sample space S.

We see the definition dictates the values of the probability of an event occurring lie within the closed interval [0,1]. If $P(E) = 0$, then the event E is called an impossibility; if $P(E) = 1$, then the event E is called a certainty. Between the values of 0 and 1, the likelihood of an event occurring is assessed by comparing the relative nearness of a value to either of those extremes. That is, events whose probabilities are closer to 0 are less likely to occur than events whose probabilities are closer to 1. An event whose probability lies squarely in the middle of this range, at 0.5, is as likely to occur as it is not to occur.

As an example, consider rolling a fair, six-sided die. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. An illustration of an event E is rolling an even number on that die; that is, $E = \{2, 4, 6\}$. The theoretical probability of rolling an even number is:

$$\mathrm{P}(E) = |E|/|S| = 3/6 = 1/2 = 0.5$$

We can express this probability either in the fractional or decimal form shown, or we can convert the result to percentages. Thus, acceptable statements of the probability in this problem are $P(E) = 0.5$, $P(E) = 1/2$, and $P(E) = 50\%$.

In the same scenario, the theoretical probability of rolling a number that is evenly divisible by 3, which we will call the event $D = \{3, 6\}$, is $2/6 = 1/3$ 0.333, or approximately 33.3%.

Notice event E is more likely to occur than event D, because $P(E) > P(D)$.

Similarly, for flipping a fair coin, the sample space is $S = \{H, T\}$. Suppose the event of getting heads is $E = \{H\}$; the probability of getting heads when flipping this coin is:

$$\mathrm{P}(E) = |E|/|S| = 1/2 = 0.5 = 50\%.$$

Recall the complement of an event E consists of those outcomes in the sample space not included in event E. For any event, an outcome is either in the event or in the complement of that event, and so the sum of the probability of an event and the probability of its complement must be 1.

That is, $P(E) + P(E^c) = 1$.

Example 11.5 In a poll of 75 residents of the small island nation of Togoland, those individuals who had a favorable view of charismatic leader Maraca Kabob numbered 31, while those who had an unfavorable view numbered 44. What is the probability a randomly selected individual from this poll would have an unfavorable view of Maraca Kabob?

Solution:

A success, in this case, is a respondent having an unfavorable view of Maraca Kabob. The number of outcomes defined as a success is 44. Let us identify U = {respondent has an unfavorable view of Maraca Kabob}.

We are aware 44 of the 75 respondents have an unfavorable view of their leader, so we calculate $P(U) = \frac{44}{75} \approx 0.59$.

Example 11.6 With a 52-card deck of cards, let E represent the event of drawing an ace from the deck. What is the complement of event E? What are the probabilities of E and E^c?

Solution:
There are four aces in the deck of 52 cards. The complement of E is the event of drawing a card that is NOT an ace from the deck.

Since there are 4 aces and 52 cards, $P(E) = \frac{4}{52} = \frac{1}{13} \approx 0.077$. There are 48 cards that are not aces, and so $P(E^c) = \frac{48}{52} = \frac{12}{13} \approx 0.923$.

Observe, once we had P(E), there was no need to calculate $P(E^c)$ as we did here. Knowing $P(E^c) = 1 - P(E)$, we could have said $P(E^c) \approx 1 - 0.077 = 0.923$.

Example 11.7 Imagine an opaque (non-transparent) box contains marbles of varying colors. The box contains 13 black marbles, 10 white marbles, 11 yellow marbles, and 6 blue marbles. What is the probability of reaching into the box and drawing a blue marble from the box, and what is the probability of drawing a marble from the box that is NOT blue?
Solution:
We will define the event B to be the act of drawing a blue marble from the box. The number of blue marbles is 6, while there are a total of 40 marbles in the box. So, $P(B) = \frac{6}{40} = \frac{3}{20} = 0.15$. The probability of NOT drawing a blue marble is $P(B^c) = 1 - 0.15 = 0.85$.

11.1.2 Empirical Probability

It is often the case that we do not know the nature of the sample space. There are multiple reasons this could be the case, but we can imagine a situation where we are drawing items from an opaque container whose contents are unknown to us (what those contents might be or how many might be within the container).

Empirical probability refers to the probability of an event occurring based on observed data or actual experience, rather than on theoretical calculations. It is also called **experimental probability** because it is derived from conducting experiments or gathering real-world data.

The definition of empirical probability is:

$$P(E) = \frac{\text{number of successes observed}}{\text{number of trials observed}}$$

Empirical probability can differ from theoretical probability, especially when the sample size is small or the trials aren't perfectly random, but as the number of trials increases, empirical probability tends to approach theoretical probability. We refer to this property as the Law of Large Numbers.

Example 11.8 A bag is known to contain a set of colored marbles, some red (R), some blue (B), and some white (W). We have no idea how many marbles are in the bag or their breakdown by color. We watch a person drawing marbles from the bag with replacement. This means the marble is drawn, its color noted, and it is returned to the bag prior to the next draw. We assume the marbles are always well-mixed.

Suppose we watch a succession of five such draws, and observe the following results:

R R B R W

What is the probability the next marble drawn from the bag is blue?

Solution:
Success has been defined as the drawing of a blue marble from the bag. We notice we have observed 1 success out of 5 trials; therefore, the empirical probability of drawing a blue marble from the bag is $\frac{1}{5} = 0.20 = 20\%$.

The potential flaw in the reasoning here is obvious: we have no idea how many marbles are in the bag, nor how many of those are blue. It is possible there are 20 marbles in the bag with just one being blue, and pure chance has produced the collection of draws we have observed. It is possible we have actually drawn the same red marble each of the times a red marble was produced (remember, the marbles were returned to the bag after each draw). In that case, the theoretical probability of drawing a blue marble from the bag is $\frac{1}{20} = 0.05 = 5\%$.

The issue at hand is that we typically do not know in advance the total number of possible outcomes in an experiment or any certainty a given outcome could be a repetition of a prior result. It is for this reason we appeal to empirical probability, understanding as we do so that we may not be getting the full picture we would have gotten had all the facts been known to us.

Example 11.9 A person regularly consults his alarm clock on waking. He notices that, on a series of 50 consecutive workdays, he has wakened prior to the alarm clock ringing on 17 separate days. What is the empirical probability the person will awaken prior to the alarm going off on any particular workday?

Solution:
The person has conducted 50 trials and observed 17 successes (viewing success as having wakened prior to the alarm clock ringing). Therefore, the empirical probability the person will awaken prior to the alarm clock ringing on any particular workday is $\frac{17}{50} = 0.34 = 34\%$.

Note it is more likely the person will not wake up prior to the alarm clock ringing than that he WILL wake up prior to the alarm clock ringing. The threshold for an "equal likelihood" of an event happening or not happening is 0.50, or 50%.

Example 11.10 Consider a cryptographic hash function, such as SHA-256. The goal of this function is to take an input (like a file, message, or password) and produce a fixed-length output (a hash value). For a well-designed cryptographic hash function, the output should appear random and uniformly distributed, meaning each possible hash value should have an equal probability of occurring.

In an empirical approach, you might perform the following experiment:

1) Collect Hash Outputs: Generate many different inputs, say 1 million unique files or messages, and apply the hash function to each. For each input, record the resulting hash value.
2) Check for Collisions: A **collision** occurs when two different inputs produce the same hash output. In a perfect hash function, collisions should be extremely rare, but if we observe the hash values of many inputs, we might see some collisions due to the finite size of the output space (e.g., a 256-bit hash has 2^{256} possible hash values, but we are only working with a finite set of inputs).

After hashing 1 million inputs, you count how many collisions occurred. If you have three collisions out of 1 million trials, what is the empirical probability of a collision occurring?

Solution:
The empirical probability of a collision is $P(collision) = \frac{\#\ of\ collisions}{\#\ of\ observed\ trials} = \frac{3}{1{,}000{,}000} = 0.000003$

In cryptography, empirical probability is useful for assessing real-world behavior, such as the likelihood of hash collisions, key collisions in symmetric encryption schemes, or vulnerabilities in cryptographic protocols that might not be obvious through theoretical analysis alone.

11.2 Axioms of Probability

In probability theory (both theoretical and empirical), the fundamental principles that govern the behavior of probabilities are known as the axioms of probability. These axioms provide a foundation for the mathematical treatment of probability and ensure the consistency and logical soundness of probability calculations.

The three axioms of probability, as introduced in 1935 by the Russian mathematician Andrey Kolmogorov (1903–1987), are as follows:

1) **Non-negativity:** For any event E in the sample space S, the probability of E is non-negative, that is, $P(E) \geq 0$.
2) **Normalization:** The probability of the entire sample space S is 1, that is, $P(S) = 1$.
3) **Additivity:** For any two mutually exclusive events E and F (i.e., events that cannot occur simultaneously), the probability of their union is the sum of their individual probabilities, that is, $P(E \cup F) = P(E) + P(F)$.

These axioms form the foundation for all probability calculations and ensure that the assigned probabilities behave in a consistent and intuitive manner.

The first and second axioms are obviously true.

For the third axiom, we recall two events E and F are mutually exclusive if they share no outcomes. The union of their sets of outcomes will, therefore, contain no duplications between the two sets of outcomes, and so $|E \cup F| = |E| + |F|$.

Interestingly, we need stronger conditions when the sample space contains infinitely many events (for instance, the amount of snowfall on a given winter day, or the price of oil at any given time). Since the cardinality of such a sample space is infinite, we can extend the Axiom of Additivity in a natural manner to include the following result: If an event $A = \{O_1, O_2, \ldots\}$ with each outcome O_i being mutually exclusive. Then, the probability of event A occurring is

$$P(A) = P\left(\bigcup_{i=1}^{\infty} O_i\right) = \sum_{i=1}^{\infty} P(O_i)$$

Establishing this result, though intuitively reasonable, lies beyond the scope of our work here and within the realm of $\sigma - algebras$. We therefore make this extension of the Axiom of Additivity as a conjecture, since we cannot prove its validity using the methods currently at our disposal.

Example 11.11 When rolling a six-sided die, the probability of getting a whole number between 1 and 6 (inclusive) is 1, as it is a certainty one of these outcomes must occur.

Example 11.12 Consider rolling a six-sided die and defining the events:

- E: Rolling an even number (2, 4, 6)
- F: Rolling a number divisible by 5 (5)

Solution:
Since these events are mutually exclusive (you cannot roll an even number and a number divisible by 5 simultaneously), the probability of rolling either an even number or a number divisible by 5 is the sum of their individual probabilities:

$P(E) = 3/6 = 1/2$
$P(F) = 1/6$
$P(E \cup F) = P(E) + P(F) = 1/2 + 1/6 = 3/6 + 1/6 = 4/6 = 2/3.$

Example 11.13 Compute the probability of rolling either a 3 or a 7 on an eight-sided die.

Solution:
There are eight distinct outcomes to the experiment, given by the sample space $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$. The events of rolling a 3 and rolling a 7 are mutually exclusive and can be represented by $E = \{3, 7\}$. The Axiom of Additivity tells us the probability of rolling either a 3 or a 7 is the sum of the individual probabilities. Assuming (as we typically do) the die is fair, each result has a 1/8 probability of occurring, and so $P(E) = 1/8 + 1/8 = 2/8 = 1/4 = 0.25 = 25\%$.

11.3 Basic Rules of Probability

11.3.1 The Addition Rule for Events That Are NOT Mutually Exclusive

When we have events that are not mutually exclusive, the formula $P(E \cup F) = P(E) + P(F)$ must be modified to account for the duplication of outcomes common to both events. That is, if we use the formula as given, we are double-counting the mutual outcomes.

If events E and F are not mutually exclusive, then $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Example 11.14 What is the probability of drawing a card from a standard deck and finding it is a face card or a diamond?

Solution:
We know there are 12 face cards (jack, queen, king) and 13 diamonds in a standard deck of cards. If we let D = card drawn is a diamond and F = card drawn is a face card, we know $P(F) = 12/52$, $P(D) = 13/52$.

Some of the diamonds, however, have been double-counted. They are the jack, queen, and king of diamonds. These form the intersection of the sets D and F, and the probability of drawing one of these three cards is 3/52.

Therefore,

a) $P(D \cup F) = P(D) + P(F) - P(D \cap F) = 13/52 + 12/52 - 3/52 = 22/52 = 11/26 \approx 0.423 = 42.3\%$

11.3.2 The Multiplication Rule for Independent Events

Two events are **independent** if the occurrence of one event does not affect the probability of the other event occurring. For example, tossing a coin and rolling a six-sided die are independent events. The result of the coin toss (Heads or Tails) has no effect on the outcome of the die roll (1 through 6).

Alternatively, two events are **dependent** if the occurrence of one event affects the probability of the other event occurring. An illustration of dependent events is drawing two cards from a deck without replacement. The probability of drawing a second specific card depends on the value of the first card, because the first draw alters the composition of the deck.

The key difference between the two is that independent events do not influence each other, while dependent events are influenced by each other in terms of probability.

If A and B are independent events, $P(A \cap B) = P(A) \cdot P(B)$.

Example 11.15 Suppose you flip a fair coin and toss a 10-sided die. What is the probability the result of the flip is a tail (T), and the result of the tossed die is divisible by 3 (D)?

Solution:
The events are clearly independent; one has no bearing on the other. We have seen $P(T) = 0.5$, and since three of the rolls of the 10-sided die are successes, $P(D) = 0.3$. Therefore,

$$P(T \cap D) = 0.5 \times 0.3 = 0.15.$$

Example 11.16 In a particular tabletop role-playing game, one situation requires a player to simultaneously roll an 8-sided die and a 20-sided die; how many different outcomes of the roll are possible?

Solution:
Though the dice are tossed simultaneously, we can arbitrarily assign the first task as the toss of the 8-sided die, and the second task as the toss of the 20-sided die. On the 8-sided die, there are 8 possible outcomes, while on the 20-sided die there are 20 possible outcomes. The total number of possible rolls of the two dice is $8 \times 20 = 160$.

Example 11.17 Suppose a restaurant offers four appetizers, six entrees, and three desserts. How many different three-course meals can be ordered?

Solution:
This problem involves three separate situational choices; though they are made in a single meal order, we can view them as a series of choices, made in order.

There are four ways to choose the appetizer. For each of the four appetizers, there are six ways to choose the entree. In turn, for each of the 24 (4×6) appetizer–entree combinations, there are three options for the dessert.

Using the multiplication principle, the total number of possible three-course meals is $4 \times 6 \times 3 = 72$.

11.3.3 Conditional Probability

The conditional probability of an event A, given that an event B has occurred, is denoted as $\mathrm{P}(A|B)$ and is defined as:

$$\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$$

where $P(A \cap B)$ is the probability of the intersection of events A and B, and $\mathrm{P}(B)$ is the probability of event B.

Example 11.18 In a bag, there are five red balls and three blue balls. If a ball is drawn from the bag and found to be red, what is the probability that the next ball drawn will also be red?

Solution:
Let A be the event of drawing a red ball on the second draw, and B be the event of drawing a red ball on the first draw.

The total number of balls in the bag is $5 + 3 = 8$.
The probability of drawing a red ball on the first draw is $P(B) = 5/8$.

The probability of drawing a red ball on the second draw, given that the first ball drawn is red, is $P(A \mid B) = (4/7)$, as there are now four red balls out of seven remaining balls in the bag.

Example 11.19 Suppose we are going to draw a card from a standard deck of 52 cards. What is the probability we draw a king, given the drawn card is a face card (jack, queen, or king).

Solution:
Recall the formula for conditional probability:

$$\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$$

The event A = draw a king, while B = draw a face card.

There are 4 kings and 12 face cards in a deck of cards, and so $P(A \cap B) = 4/52$, while $P(B) = 12/52$. Therefore,

$$\begin{aligned}\mathrm{P}(A \mid B) &= \mathrm{P}(A \cap B)/\mathrm{P}(B)\\ &= \frac{\frac{4}{52}}{\frac{12}{52}}\\ &= \frac{4}{12}\\ &\approx 0.333\\ &\approx 33.3\%\end{aligned}$$

11.3.4 The Multiplication Rule for Dependent Events

When two events A and B are dependent, the probability of them both occurring is the product of the probability of A occurring and the conditional probability of B occurring, given that A has already occurred. We express the probability that B occurs, given A has already occurred as $P(B \mid A)$.

That is,

$$P(A \cap B) = P(A) \cdot P(B \mid A).$$

Example 11.20 Suppose you have a deck of 52 playing cards, and you draw two cards without replacement. Let event A represent drawing a red card (either a heart or a diamond) and event B represent drawing a red card on the second draw. What is the probability of drawing a red card on both draws?

Solution:
On the first draw, $P(A) = 26/52 = 1/2$, while on the second draw, $P(B \mid A) = 25/51$. The reason for the change in probability for the second draw is that (after the first card has been drawn) there are only 25 red cards remaining among the 51 cards yet to be dealt.

Therefore,

$$P(A \cap B) = \frac{1}{2} \cdot \frac{25}{51} = \frac{25}{102} \approx 0.245.$$

Example 11.21 Consider a scenario in which a computer network consists of two servers, A and B, and we are interested in the probability that both servers are functioning at the same time. Let Event 1 represent the situation where server A is working, and $P(A) = 0.95$ (i.e., there is a 95%

chance that server A is working). Let Event 2 represent the situation where server B is working, given that server A is working, and $P(B \mid A) = 0.9$ (i.e., if server A is working, there's a 90% chance that server B will also be working).

What is the probability both servers are working?

Solution:
The events are dependent because the functioning of server B depends on whether server A is functioning or not. According to the multiplication rule for dependent events, the probability that both events occur (i.e., both servers A and B are working) is:

$$P(A \cap B) = 0.95 \cdot 0.90 = 0.855$$

So, the probability that both servers are working at the same time is 0.855 or 85.5%.

This type of calculation is important in network reliability analysis, consideration of fault tolerance in distributed systems, or when modeling probabilistic behavior in software systems.

Example 11.22 In a factory, 80% of the products are inspected, and 95% of the inspected products are deemed acceptable. What is the probability that a randomly selected product is both inspected and acceptable?

Solution:
Let A be the event that a product is inspected, and B be the event that a product is acceptable.

$P(A) = 0.8$ (80% of the products are inspected)

$P(B|A) = 0.95$ (95% of the inspected products are acceptable)Using the multiplication rule, the probability that a product is both inspected and acceptable is:

$P(A \cap B) = P(B|A) \times P(A) = 0.95 \times 0.8 = 0.76$.

11.4 Counting Techniques and Probability

One of the challenges faced in probability theory involves counting the number of possible outcomes for a given experiment. To determine the theoretical probability of an event, we need to know the total number of possible outcomes of the experiment and the number of outcomes for that event.

Two important notions with which we must be familiar are permutations and combinations.

11.4.1 Permutations and Combinations

The multiplication principle presumes we are searching for the number of all possible alternatives when a series of choices (each having its own specific number of options) are to be made. We contrast that notion to permutations and combinations.

Permutations and combinations are methods for counting outcomes when the order of selections matters or does not matter, respectively.

When stating the order is "relevant" or "irrelevant," we mean that a reorganization of the members of the group (or the stated ordering of the members) does or does not affect the nature of the group.

Consider the creation of a subcommittee from a group of individuals serving on an overarching super-committee. One such subcommittee could be three individuals charged with gathering

information. No specific roles are assigned to any person; they are simply members of the group. The order in which the members are listed is not relevant.

A second subcommittee could be created where three individuals are identified as President, Vice President, and Secretary of the overall committee. The order in which the individuals are identified dictates their role. Thus, if the members are stated as Ihrain, Jose, and Katelyn, Ihrain is the President, Jose the Vice President, and Katelyn the Secretary. On the other hand, the listing of Ihrain, Katelyn, and Jose would have the roles changed: Ihrain is now the President, Katelyn the Vice President, and Jose the Secretary.

A selection in which some number of members is chosen from an overall group where the order of selection matters is called a **permutation**. A selection in which some number of members is chosen from an overall group where the order of selection does not matter is called a **combination**.

11.4.2 Permutations

Permutations are the number of ways to arrange a set of distinct objects in a specific order. The formula for the number of permutations of n distinct objects taken r at a time is:

$$\mathrm{P}(n, r) = \frac{n!}{(n-r)!}$$

where $n!$ (read as "n-factorial") is the product of all positive integers less than or equal to n. To prevent contradictions and handle computational needs, we specifically define 0! to be equal to 1.

Example 11.23 In how many ways can five people be seated in a row?

Solution:
There are five people, and we want to arrange them in a row. If the individuals are Adam (A), Barb (B), Celine (C), Dao (D), and Ephraim (E), the arrangement of individuals seated from left to right as ABCDE is NOT the same as the arrangement ABDEC. Therefore, rearrangements of the individuals is a permutation of the group.

The number of permutations of the five individuals is given by

$$\mathrm{P}(5, 5) = 5\,!\,/(5-5)\,!\,= 5\,!\,/0\,!\,= 5\,!\,= 120$$

Example 11.24 In a committee of 20 individuals, three are selected randomly to sit on a special leadership subcommittee. The first person chosen will be the President, the second will be the Vice President, and the third will be the Secretary. How many possible groups of three people can be chosen from the overall group of 20?

Solution:
Because the order of selection is relevant in this scenario, we are considering a permutation. The size of the overall group is $n = 20$, and the number of individuals being chosen is $r = 3$. Therefore,

$$\mathrm{P}(20, 3) = \frac{20!}{17!}$$

Using the definition of factorials, we do a partial simplification of 20!:

$$\mathrm{P}(20, 3) = \frac{(20)(19)(18)(17!)}{(17!)}$$

The reason for the partial simplification is that we know the remainder of the expansion of 20! (from 17 down to 1) will reduce by 17! from the denominator. Continuing,

$$P(20, 3) = \frac{6840}{1} = 6840$$

There are 6840 possible ways to select the three-person leadership subcommittee.

Example 11.25 Numbers are usually given in "decimal" form, the form with which we are most familiar. There are other representation forms for numbers, such as binary, trinary, octal, hexadecimal, and so forth. Without going into too much of the theory here, an octal number (as an illustration) only uses the digits 0, 1, 2, 3, 4, 5, 6, 7.

Suppose a number expressed in octal form consists of three distinct nonzero digits. How many such octal numbers are possible?

Solution:
Changing the order of the digits changes the value of the number, so this is a case where the order of the numbers matters. Thus, we are asking for the number of permutations of three digits chosen from a set of seven possible digits. The number of such permutations is

$$P(7, 3) = \frac{7!}{4!} = \frac{(7)(6)(5)(4!)}{4!} = \frac{(7)(6)(5)}{1}$$

$$P(7, 3) = 210$$

11.4.3 Combinations

Combinations are selections of a subset of r objects from a set of n distinct objects, where the order of the objects does not matter. The formula for the number of combinations of n distinct objects taken r at a time is:

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

Example 11.26 How many ways can three books be chosen from a bookshelf holding eight books?

Solution:
There are three books, from which we want to choose three. The number of combinations is given by

$$C(8, 3) = \frac{8!}{3!(8 - 3)!} = \frac{8!}{3!5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{3 \cdot 2 \cdot 1 \cdot 5!} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} = 56$$

Example 11.27 In a committee of 20 individuals, three are selected randomly to sit on a fact-finding subcommittee. How many possible groups of three people can be chosen from the overall group of 20?

Solution:
Because the order of selection is irrelevant in this scenario, we are considering a combination. The size of the overall group is $n = 20$, the number of individuals being chosen is $r = 3$. Therefore,

$$C(20, 3) = \frac{20!}{3!17!}$$

$$C(20, 3) = \frac{(20)(19)(18)(17!)}{(3)(2)(1)(17!)}$$

The 17! factors will reduce, and observe that 3 will reduce with the 18, and that 2 will reduce with the 20. Therefore,

$$C(20, 3) = \frac{(10)(19)(6)}{(1)(1)(1)} = 1140$$

11.5 Random Variables

A variable is called "random" if it can take on different values, each with a certain probability of occurrence. Random variables can be classified as either discrete or continuous, depending on the nature of the values they can take.

Discrete random variables can take on a countable number of distinct values, such as integers or a finite set of values.

> Note: The concept of "countability" is somewhat abstract. Technically, it refers to a set that is either finite or that can be put into a one-to-one correspondence with the natural numbers. A less formal definition is that a set is countable if we can find the cardinality (number of members) of the set.

Example 11.28 The number of heads obtained when a fair coin is tossed five times is a discrete random variable, as it can take on the values 0, 1, 2, 3, 4, or 5.

Continuous random variables are random variables that can take any value within a certain range or interval. The key characteristic of a continuous random variable is that it can assume an infinite number of possible values, usually represented by real numbers. These values are not discrete, meaning they are not limited to specific, countable points, but can instead take on any value within a given range.

For example, consider a random variable that represents the time it takes for a bus to arrive at a given station. The time could be any real number, such as 5.3 minutes, 5.34 minutes, 5.345 minutes, and so on, with an infinite number of possible outcomes within a certain range (e.g., between 0 and 10 minutes).

The temperature in a room is another example of a continuous random variable. If we know the temperature in the room ranges from 70 to 72°F on a given day, then at any given moment, the temperature could be any of the infinitely many real number values between 70 and 72°F, inclusive.

11.6 Probability Mass Function

The probability mass function (PMF) of a discrete random variable X is a function that gives the probability that X takes on each of its possible values. The PMF is denoted as $P(X = x)$, where x represents the possible values of the random variable X.

The PMF must satisfy the following properties:

1) $P(X = x) \geq 0$ for all possible values of x.
2) The sum of the probabilities of all possible values of X must be equal to 1.

Example 11.29 Consider a discrete random variable X that represents the number of heads obtained when a fair coin is tossed three times. The sample space has eight members:

{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

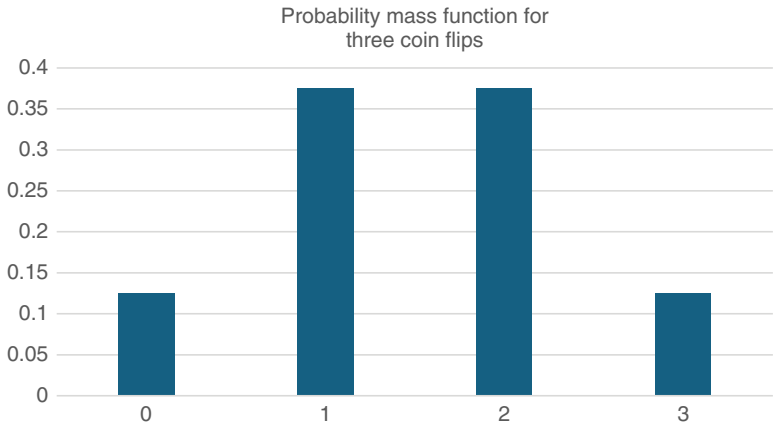
The possible values of X are 0, 1, 2, and 3, and the corresponding PMF is:

X	$P(X = x)$
0	0.125
1	0.375
2	0.375
3	0.125

The probabilities in the table above can be calculated using the binomial distribution formula or by referring to the sample space and the number of successes meeting the criteria for each value of X.

11.6.1 Graphical Representation of the PMF

The PMF of a discrete random variable can be represented graphically using a bar chart or a histogram, where the height of each bar corresponds to the probability of the associated value of the random variable.



In this graph, the x-axis represents the possible values of the random variable, and the y-axis represents the corresponding probabilities.

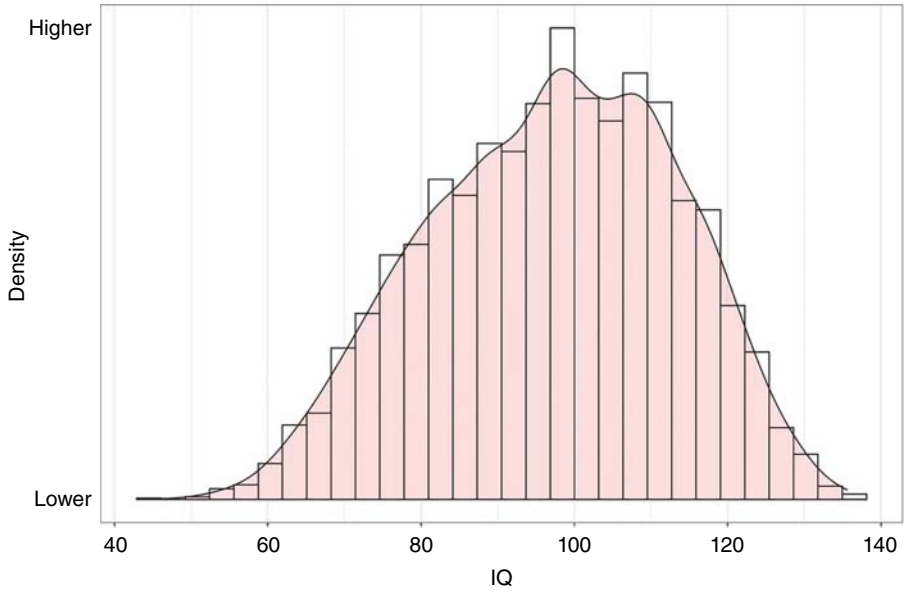
11.7 Distributions of Discrete Data

Data can be "distributed" in a number of different ways; before examining several examples, it is worth while to elaborate upon what is meant by this concept.

Imagine a situation where 10 people take a particular test, and their grades are A, A, B, B, C, C, D, D, F, F. For each grade level, we see there are exactly two individuals obtaining that grade.

That is, each grade level has a uniform number of students obtaining that grade—we refer to this as a "uniform" distribution of data (naturally, in a real-life situation, it is unlikely there would be a perfectly uniform distribution, but rather simply an approximately uniform distribution).

As a second example, consider the "IQ test" results for a large number of individuals. Scores on IQ tests generally follow a "bell curve" distribution (a symmetric distribution of scores centered around the mean). We call this a "normal" distribution. Here is a graph of a normal distribution of a sampling of IQ scores:



Note the distribution is not perfectly bell-shaped, but for a sampling of individuals, we would not expect that real-world data to conform precisely to a particular shape.

We will now turn our attention to other forms of data distribution, including examples that follow those distributions.

11.7.1 Bernoulli Distribution

The Bernoulli distribution, first described by Jacob Bernoulli (1655–1705), is a discrete probability distribution that models the outcome of a single trial with two possible outcomes, typically referred to as "success" and "failure." The random variable X in a Bernoulli distribution can take the values 0 (failure) or 1 (success).

The PMF of a Bernoulli random variable X is given by:

$$P(X = x) = p^x * (1 - p)^{(1-x)}, \text{ where } x = 0 \text{ or } 1$$

Here, p is the probability of success (the event of interest) in a single trial.

Example 11.30 A fair coin is tossed, and the random variable X represents the outcome, where $X = 1$ if the outcome is "heads" and $X = 0$ if the outcome is "tails." The PMF of X is:

$$P(X = 1) = 0.5$$
$$P(X = 0) = 0.5$$

11.7.2 Binomial Distribution

The binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of independent Bernoulli trials, where each trial has only two possible outcomes (success or failure).

The PMF of a binomial random variable X with n trials and probability of success p is given by:

$$\mathrm{P}(X = x) = \binom{n}{x} * p^x * (1-p)^{(n-x)}, \quad \text{where } x = 0, 1, 2, \ldots, n$$

Here, $\binom{n}{x}$ represents the binomial coefficient, which is the number of ways to choose x successes out of n trials.

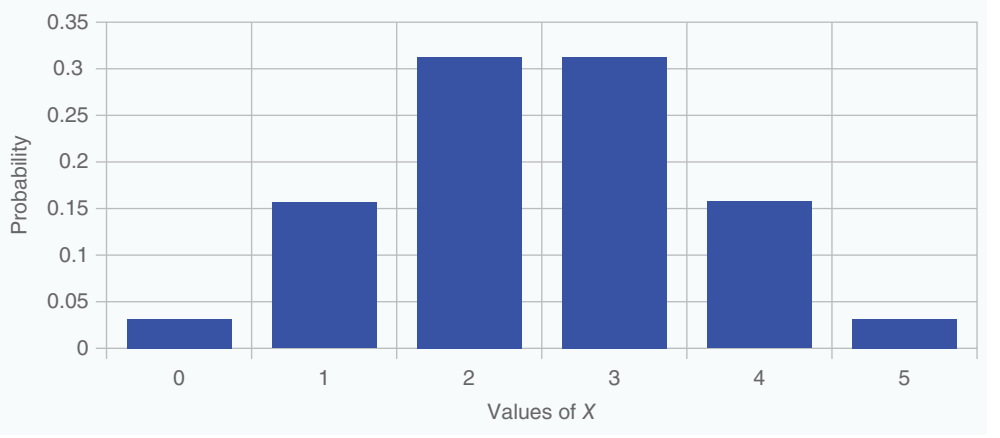
Example 11.31 A fair coin is tossed five times. The random variable X represents the number of heads obtained. The PMF of X is:

X	P(X = x)
0	0.0312
1	0.1562
2	0.3125
3	0.3125
4	0.1562
5	0.0312

As an illustration of how the values on the table are found will be instructive at this point. We will compute the value for $X = 1$ and leave the rest for the reader.

$$\mathrm{P}(\mathrm{X} = 1) = \binom{5}{1}(0.5)^1(0.5)^4 = \frac{5!}{1!4!}(0.5)(0.0625) = 0.15625$$

This distribution can be represented graphically as follows:



11.7.3 Geometric Distribution

The geometric distribution is a discrete probability distribution that models the number of Bernoulli trials required to obtain the first success, where each trial has a fixed probability of success.

The PMF of a geometric random variable X with probability of success p is given by:

$$P(X = x) = p * (1 - p)^{(x-1)}, \text{ where } x = 1, 2, 3, \ldots$$

Example 11.32 A fair coin is repeatedly tossed until the first "heads" is obtained. The random variable X represents the number of tosses required to obtain the first "heads." The PMF of X is:

X	P(X = x)
1	0.5
2	0.25
3	0.125
4	0.0625
5	0.03125

Example 11.33 Suppose the probability of a manufacturing process producing a defective item is 0.0185; if an inspector pulls a succession of items from the assembly line, what is the geometric distribution of the number of trials needed to find a defective item?

Solution:
The random variable X represents the number of items needed to find a defective product. Let us first compute $P(X = 1)$, the probability the first item pulled from the line is defective.

$$P(X = 1) = (0.0185)(0.9815)^0 = 0.0185.$$

This is not surprising, on its face. We know that the probability of a randomly selected item being defective is 0.0185, so this seems obvious.

Now let us compute $P(X = 2)$, the probability that two trials are required to find a defective product.

$$P(X = 2) = (0.0185)(0.9815) = 0.01816$$

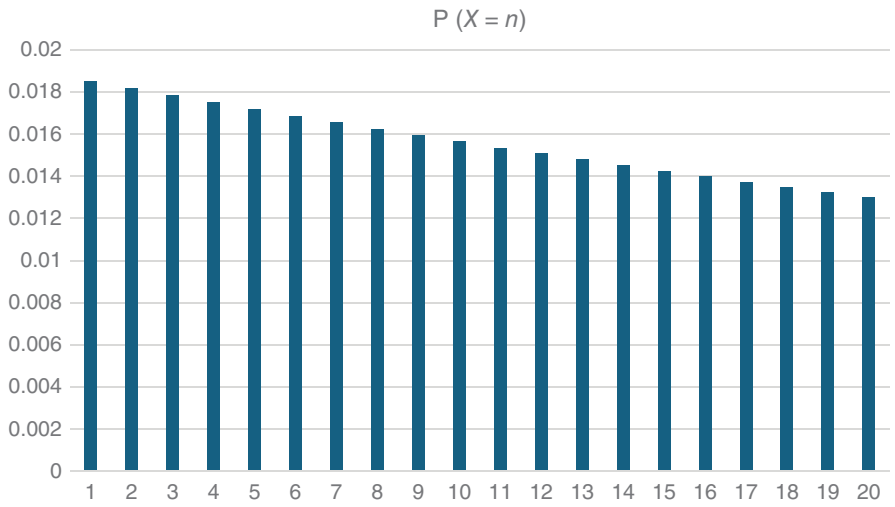
There is a 1.816% probability the second item pulled from the line is the first defective unit found.

Similarly, $P(X = 3) = 0.01782$, $P(X = 4) = 0.01749$, and so on. A brief table of these values follows:

n	P(*X* = *n*)
1	0.0185
2	0.01815775
3	0.01782183
4	0.01749213
5	0.01716852
6	0.01685091
7	0.01653916
8	0.01623319
9	0.01593288
10	0.01563812
11	0.01534881
12	0.01506486
13	0.01478616

n	P($X = n$)
14	0.01451262
15	0.01424413
16	0.01398062
17	0.01372197
18	0.01346812
19	0.01321896
20	0.01297441

If we consider the graph of the probability mass function for this situation, we find:



Carefully note the vertical scale of the chart.

An interesting observation is made by calculating the sum of the first twenty entries of the probabilities (approximately 0.312). This tells us the probability of the first defective unit being among the first 20 pulled from the line is 31.2%.

11.7.4 Poisson Distribution

The Poisson distribution (named after Baron Siméon Denis Poisson, 1781–1840) is a discrete probability distribution that models the number of events occurring in a fixed interval of time or space, given that the events occur with a constant average rate and independently of the time since the last event.

The PMF of a Poisson random variable X with parameter λ (the average number of events per interval) is given by:

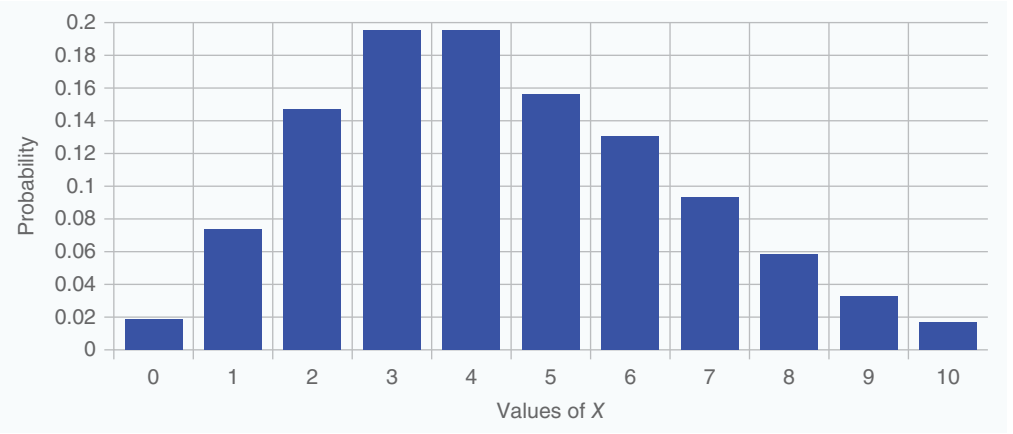
$$P(X = x) = (e^{(-\lambda)} * \lambda^x)/x!, \quad \text{where } x = 0, 1, 2, 3, \ldots$$

Here, e is the base of the natural logarithm, approximately equal to 2.71828.

Example 11.34 The average number of calls received by a customer service hotline in a 30-minute period is 4. The random variable X represents the number of calls received in a given 30-minute period. The PMF of X is:

X	P(X = x)
0	0.0183
1	0.0733
2	0.1465
3	0.1954
4	0.1954
5	0.1563
6	0.1302
7	0.0930
8	0.0581
9	0.0326
10	0.0163

This distribution can be represented graphically as follows:



Example 11.35 Website hosting companies use the Poisson distribution to model the number of expected visitors per hour that particular websites will receive. This is significant because it gives hosting companies an idea of how much bandwidth to provide to different websites to ensure that they can manage a particular number of visitors each hour.

Suppose a given site receives an average of 29 visitors per hour. What is the probability the site will experience 40 visitors in any given hour of the day?

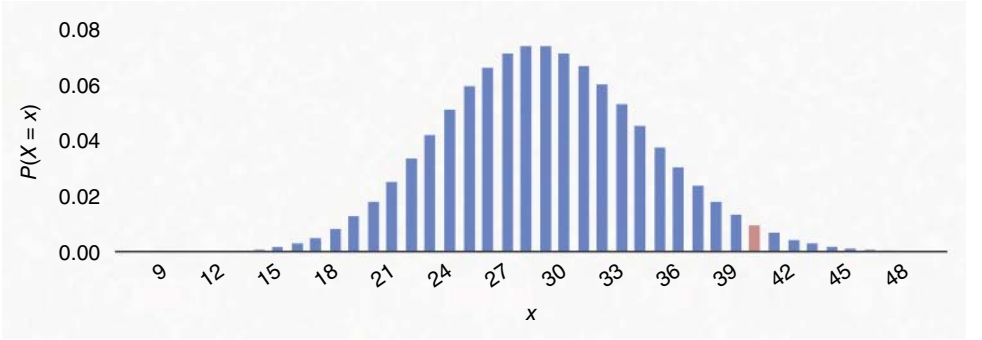
Solution:
Using the formula for the Poisson Distribution, we see

$$P(X = 40) = (e^{(-29)} * 29^{40})/40! \approx 0.00977.$$

The numbers involved here are clearly quite large and quite small. 40!, for example, is 815,915,283,247,897,734,345,611,269,596,115,894,272,000,000,000, while

$e^{-29} = 0.000000000000254366564737692$, and
$29^{40} = 31{,}327{,}079{,}986{,}911{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000{,}000$

Clearly, without computer assistance, the calculation would be daunting, to say the least! Again, using computer assistance, we can graph the distribution for this example:



The red rectangle shown is the probability for 40 visitors to the site in an hour. Is it surprising to you the highest rectangle occurs at $x = 29$?

11.7.5 Hypergeometric Distribution

The hypergeometric distribution is a discrete probability distribution that models the number of successes in a fixed number of draws, without replacement, from a finite population that contains a known number of successes.

The PMF of a hypergeometric random variable X with population size N, number of successes in the population K, and number of draws n is given by:

$$\mathrm{P}(X = x) = ((K \text{ choose } x) * (N - K \text{ choose } n - x))/(N \text{ choose } n),$$
$$\text{where } \; x = 0, 1, 2, \ldots, \min(K, n)$$

Here, (K choose x) represents the binomial coefficient, which is the number of ways to choose x successes out of K items.

The hypergeometric distribution is useful in situations where the probability of success in a draw depends on the number of successes already drawn, such as in quality control sampling, survey sampling, and certain types of card games.

Example 11.36 Suppose a beer manufacturer is concerned about the fluid content of bottles being prepared for distribution. Historically, the rate at which bottles are properly filled is 98%. A sample of 20 bottles is taken from the line without replacement from a batch of 300 bottles. What is the probability exactly 19 bottles will have content within acceptable standards?

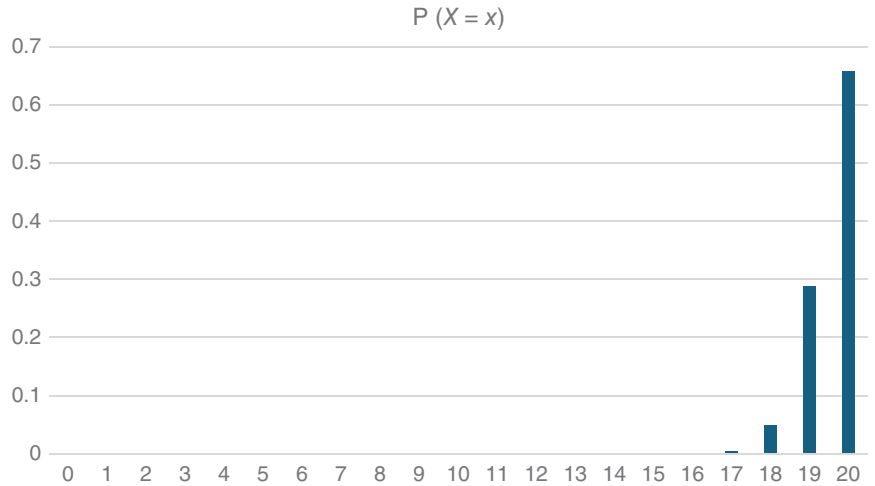
Solution:
We are interested in $P(X = 19)$, where the size of the population is $N = 300$, n is the number of successes (here, 19), and K is the historical number of properly filled bottles out of 300 (here, 294), and $x = 19$. We compute,

$$\mathrm{P}(X = 19) = \frac{\binom{294}{19}\binom{6}{1}}{\binom{300}{20}} = 0.2874076$$

We can work out the values for all possible values of x,

x	$P(X = x)$
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	4.0257E−08
15	4.5087E−06
16	0.00019655
17	0.00428563
18	0.04946326
19	0.2874076
20	0.65864241

The graph of the overall distribution for a sample of size 20 is:



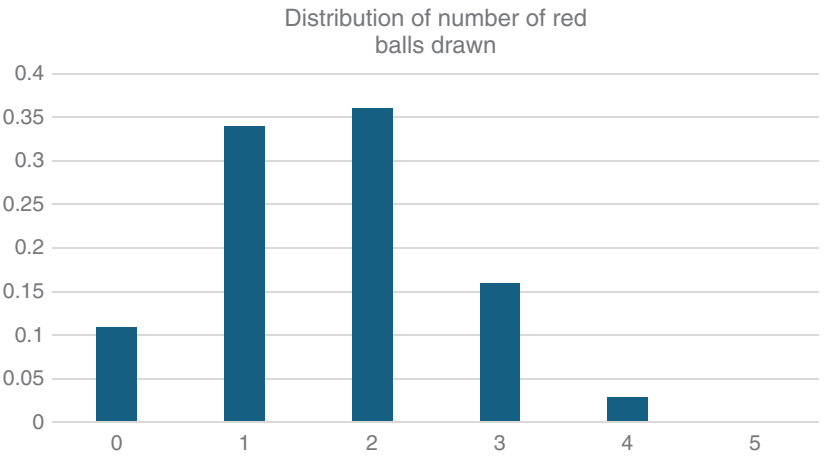
An interesting observation regarding the results of the computations in the hypergeometric distribution is found when considering the sum of the probabilities in the table; that sum is always 1.

Example 11.37 A bag contains 10 red balls and 20 blue balls. Five balls are drawn from the bag without replacement. The random variable X represents the number of red balls drawn. The PMF of X is:

X	P(X = x)
0	0.2077
1	0.4154
2	0.2769
3	0.0923
4	0.0077
5	0.0000

x	P(X = x)
0	0.10879542
1	0.33998568
2	0.35998484
3	0.15999326
4	0.02947244
5	0.00176835

This distribution can be represented graphically as follows:



11.7.6 Expected Value and Variance of Discrete Random Variables

The expected value (or mean) of a discrete random variable X is the long-term average or central tendency of the distribution. It is denoted by E(X) or μ. As an illustration, it is known the IQ test

is a "standardized" dataset, with a mean value of 100 (this is "by design"; that is, the test scores are scaled in such a way that the mean will be equal to 100, which is generally described as "average intelligence"). Therefore, if we were to take larger and larger groups of individuals and subject them to an IQ test, the mean would tend toward a value of 100 as the samples became larger. That is, we EXPECT the mean of ever-increasing samples to tend toward the value of 100.

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$ and corresponding probabilities $p_1, p_2, \ldots, p_n$, the expected value is calculated as:

$$\mathrm{E}(X) = \Sigma(xi * p_i), \quad \text{where } i = 1, 2, \ldots, n$$

Example 11.38 Consider the random variable X representing the number of heads obtained when a fair coin is tossed five times. The possible values of X and their corresponding probabilities are:

X	P(X = x)
0	0.0312
1	0.1562
2	0.3125
3	0.3125
4	0.1562
5	0.0312

The expected value of X is:

$$\mathrm{E}(X) = 0 * 0.0312 + 1 * 0.1562 + 2 * 0.3125 + 3 * 0.3125 + 4 * 0.1562 + 5 * 0.0312 = 2.5$$

Example 11.39 Imagine we have a six-sided die printed in such a way that three of the sides have just one pip, one side has two pips, and the other side has four pips. If the die is rolled repeatedly, what is the expected value of any roll of this die?

Solution:
The first thing to do is to assemble the list of potential outcomes and their probabilities.

die result	probability
1	1/2
2	1/6
4	1/6

Next, we calculate the sum

$$\sum_{i=1}^{3} x_i p_i = (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{6}\right) + (4)\left(\frac{1}{6}\right) = \frac{1}{2} + \frac{1}{3} + \frac{2}{3} = \frac{9}{6} = \frac{3}{2}$$

The calculation tells us that, if we continued to roll the die and track its total, the long-term average result of the roll would be $\frac{3}{2}$.

It is immaterial that this average is not one of the face values of the die! Remember, we are considering the average of all the rolls, and it is frequently the case an average value might never be attained by any individual trial.

11.7.7 Variance

The variance of a discrete random variable X measures the spread or dispersion of the distribution around its expected value. It is denoted by $\text{Var}(X)$ or σ^2.

The formula for the variance of a discrete random variable X is:

$$\text{Var}(X) = \Sigma((xi - \mu)^2 * \text{pi}), \text{ where } i = 1, 2, \ldots, n, \text{ and } \mu = \text{E}(X)$$

The square root of the variance, denoted by σ, is called the standard deviation, which provides a measure of the typical deviation from the expected value.

Example 11.40 Continuing with the previous example, let's calculate the variance of the random variable X representing the number of heads obtained when a fair coin is tossed five times.

The expected value $E(X) = 2.5$, as calculated earlier.

$$\begin{aligned}\text{Var}(X) &= (0 - 2.5)^2 * 0.0312 + (1 - 2.5)^2 * 0.1562 + (2 - 2.5)^2 * 0.3125 \\ &\quad + (3 - 2.5)^2 * 0.3125 + (4 - 2.5)^2 * 0.1562 + (5 - 2.5)^2 * 0.0312 \\ &= 6.25\end{aligned}$$

The standard deviation is the square root of the variance, which is $\sqrt{6.25} = 2.5$.

Why do we seem to have two measures of dispersion for the discrete random variable X? It seems like one ought to suffice, yet there are two. It is worthwhile to recall the distinction between and the interpretation of each measure.

Variance is the average of the squared differences from the mean, while standard deviation is the square root of the variance. This fundamental distinction has implications for how you interpret and use these measures in your analysis.

The variance provides a figure that expresses how widely dispersed the data points are around the mean. However, because variance uses squared units, it can be difficult to interpret in the context of the original data.

Standard deviation, which takes the square root of variance, yields the original units of the data. This makes it easier to understand and interpret because it provides a measure of spread that is directly comparable to the mean of the dataset. Standard deviation is widely used because it offers a clear picture of how much individual data points deviate from the mean.

Example 11.41 Suppose two shops X and Y have four employees each, and we wish to compare the rates of compensation between the two shops. In shop X, two employees earn \$14 per hour and the other two earn \$16 per hour. In shop Y, one employee earns \$11 per hour, one earns \$10 per hour, the third earns \$19, and the fourth receives \$20 per hour. The average hourly wage for each shop is \$15 but are the employees in the two shops being compensated in a similar manner?

Solution:

We calculate the standard deviation of the datasets:

Employee	Shop *X*	Shop *Y*
1	\$14.00	\$11.00
2	\$14.00	\$10.00
3	\$16.00	\$19.00
4	\$16.00	\$20.00
Std Dev	?	?

For shop X, we compute the variance (the square root of which will be the standard deviation),

$$\sum_{i=1}^{4}(x_i - \mu)^2 p_i = (14-15)^2(0.25) + (14-15)^2(0.25) + (16-15)^2(0.25) + (16-15)^2(0.25) = 1$$

Clearly, the square root of the variance is 1, so the standard deviation is 1. Note that, in this case, the result is obvious, since all the data values differ from the mean by \$1; that is the nature of the standard deviation: it is the typical distance of any data point from the mean.

The process is repeated for Shop Y:

$$\begin{aligned}\sum_{i=1}^{4}(x_i - \mu)^2 p_i &= (11-15)^2(0.25) + (10-15)^2(0.25) + (19-15)^2(0.25) + (20-15)^2(0.25) \\ &= (16)(0.25) + (25)(0.25) + (16)(0.25) + (25)(0.25) \\ &= 4 + 6.25 + 4 + 6.25 \\ &= 20.5\end{aligned}$$

The square root of the variance is approximately \$4.53.

Remember, the standard deviation describes the spread of the data from the mean; the data with the larger standard deviation is more spread out than the data with the smaller standard deviation. Thus, though the mean salaries at both shops are identical, the salaries at Shop Y are spread apart significantly more than those at Shop X.

11.8 Continuous Probability Distributions

11.8.1 Probability Density Functions

Unlike the PMF used for discrete random variables, the probability density function (PDF) does not give the actual probability of a specific value. Instead, it gives the probability density, which can be used to calculate the probability of a random variable falling within a given interval.

The key properties of a probability density function f(x) are:

- f(x) ≥ 0 for all x in the domain of the random variable.
- The total area under the curve of f(x) is equal to 1.

To calculate the probability that a continuous random variable X falls within an interval $[a, b]$:

- Divide the interval $[a, b]$ into a large number of small subintervals.
- For each subinterval, calculate the width (length) of the subinterval and the average height of the PDF within that subinterval.
- Multiply the width and the average height to get the approximate area of the subinterval.
- Sum the areas of all the subintervals to get an approximate value for the probability $P(a \leq X \leq b)$.
- As the number of subintervals increases, the approximate probability approaches the exact probability.

> Note: The second bullet point may appear to be casually stated, but a certain amount of integral calculus lies behind it. Recall the average value of a continuous function f(x) over

a closed interval [a,b] is found using $\frac{1}{b-a}\int_a^b f(x)dx$. We will demonstrate this for the first example and thereafter leave the computation to the reader.

Example 11.42 The lifespan of a certain type of light bulb is modeled by a continuous random variable X with the following probability density function:

$$f(x) = 0.5 * e^{(-0.5x)}, \text{ for } x \geq 0$$

Here, the random variable X represents the lifespan of the light bulb in years. What is the probability a randomly selected light bulb will have a lifespan between four and eight years?

Solution:
To calculate the probability that a light bulb's lifespan is between 4 and 8 years, we can divide the interval [4, 8] into 40 subintervals, each of width 0.1 years.

For the first subinterval, we compute the average height:

Subinterval	Average Height of f(x)	Area of Subinterval
[4.0, 4.1]	0.0660	0.0066
[4.1, 4.2]	0.0627	0.00627
[4.2, 4.3]	0.0597	0.00597
…	…	…
[7.9, 8.0]	0.0093	0.00093

Summing the areas of all the subintervals, we get an approximate value for the probability:

$$P(4 \leq X \leq 8) \approx 0.117$$

Thus, the probability that one such life bulb has a life span between 4 and 8 years is approximately 0.117, or 11.7%.

11.8.2 Distributions of Continuous Random Variables

There are several important distributions that are widely used in various applications. Here, we will explore three of the most common distributions of continuous random variables: the Uniform, Normal, and Exponential distributions.

11.8.3 Uniform Distribution

The Uniform Distribution is a continuous probability distribution where the random variable X can take on any value within a specified interval $[a, b]$ with equal probability. The probability density function for the Uniform Distribution is:

$$f(x) = 1/(b-a), \text{ for } a \leq x \leq b$$

Example 11.43 Suppose the height of adult individuals in a certain population is modeled by a Uniform Distribution with $a = 160$ cm and $b = 190$ cm.

To calculate the probability that a randomly selected individual has a height between 170 cm and 180 cm, we can divide the interval [170, 180] into multiple subintervals:

Subinterval	Width of Subinterval	Average Height of f(*x*)	Area of Subinterval
[170, 171]	1 cm	0.033	0.033
[171, 172]	1 cm	0.033	0.033
[172, 173]	1 cm	0.033	0.033
…	…	…	…
[179, 180]	1 cm	0.033	0.033

Summing the areas of all the subintervals, we get an approximate value for the probability:

$$P(170 \leq X \leq 180) \approx 0.33, \text{ or } 33\%.$$

As the number of subintervals increases, the approximate probability approaches the exact probability, which is $\frac{1}{3}$ (calculated as (180–170)/(190–160)).

11.8.4 Normal Distribution

The Normal Distribution, also known as the Gaussian Distribution, is a widely used continuous probability distribution. It is characterized by a symmetric, bell-shaped curve and is defined by two parameters: the mean (μ) and the standard deviation (σ). The reason for the nomenclature "normal" is that most data in the "real world" follows the normal distribution; that is, it is the distribution we normally encounter.

> Note: The notion of a bell curve has, historically, come under attack as being problematic within social constructs. In 1994, Richard Herrnstein and Charles Murray used normal distributions in their controversial work, "*The Bell Curve: Intelligence and Class Structure in American Life*." They posited a caste structure existed in the United States, attributable to inherent societal qualities which could be described by normal distributions. It is beyond the scope of this work to further examine the controversy, but it does illustrate how the use of statistical information can (for good or ill) arise in the social sciences and public policy.

The probability density function for the Normal Distribution is:

$$f(x) = (1/(\sigma * \sqrt{(2\pi)})) * e^{(-(x-\mu)^2/(2*\sigma^2))}$$

Example 11.44 Suppose the weight of adult individuals in a certain population is modeled by a Normal Distribution with $\mu = 70$ kg and $\sigma = 5$ kg.

To calculate the probability that a randomly selected individual weighs between 65 kg and 75 kg, we can divide the interval [65, 75] into multiple subintervals:

Subinterval	Width of Subinterval	Average Height of f(*x*)	Area of Subinterval
[65, 66]	1 kg	0.0400	0.0400
[66, 67]	1 kg	0.0484	0.0484
[67, 68]	1 kg	0.0579	0.0579
…	…	…	…
[74, 75]	1 kg	0.0484	0.0484

Summing the areas of all the subintervals, we get an approximate value for the probability:

$$P(65 \leq X \leq 75) \approx 0.6826$$

As the number of subintervals increases, the approximate probability approaches the exact probability, which is 0.6826 (calculated using the standard Normal distribution table).

> Note: Calculation of probabilities such as this is more efficiently performed using the standard Normal distribution tables presented in a course in statistics. Copious amounts of time have been spent compiling such tables, and it would be misleading to suggest the tabular method of approximation given above would be a more efficient approach.

11.8.5 Exponential Distribution

The Exponential Distribution is a continuous probability distribution that models the time between events in a Poisson process, where events occur at a constant average rate. The probability density function for the Exponential Distribution is:

$$f(x) = \lambda e^{-\lambda x}, \text{ for } x \geq 0$$

where λ is the rate parameter, which represents the average number of events per unit of time.

Example 11.45 Suppose the time between customer arrivals at a bank is modeled by an Exponential Distribution with $\lambda = 0.2$ customers per minute. What is the probability the elapsed time between the arrival of two bank customers is between 2 and 5 minutes?

Solution:
To calculate the probability that the time between two customer arrivals is between 2 minutes and 5 minutes, we can divide the interval [2, 5] into multiple subintervals:

Subinterval	Width of Subinterval	Average Height of f(*x*)	Area of Subinterval
[2, 2.5]	0.5 minutes	0.0905	0.0453
[2.5, 3]	0.5 minutes	0.0826	0.0413
[3, 3.5]	0.5 minutes	0.0755	0.0377
…	…	…	…
[4.5, 5]	0.5 minutes	0.0553	0.0277

Summing the areas of all the subintervals, we get an approximate value for the probability:

$$P(2 \leq X \leq 5) \approx 0.1520$$

As the number of subintervals increases, the approximate probability approaches the exact probability, which is 0.1516 (calculated using the cumulative distribution function of the Exponential Distribution).

Using calculus, we can formalize the notion of "the number of subintervals increasing" and conclude that to calculate probabilities for specific probability density functions, the cumulative density function (cdf) is used. The cdf is the integral of the pdf and is:

$$\int_0^{\infty} \frac{1}{\mu} e^{-\frac{x}{\mu}} dx = 1 - e^{-\frac{x}{\mu}}$$

Values for an exponential random variable occur in the following way: there are fewer large values and more small values. For example, studies have shown that the amount of money customers spend in one trip to the supermarket follows an exponential distribution. There are more people who spend small amounts of money and fewer people who spend large amounts of money. If we pause to reflect, on the face of it, this is eminently reasonable.

Exponential distributions are commonly used in calculations of product reliability or the length of time a product lasts.

Example 11.46 The number of days prior to departure travelers purchase their airline tickets can be modeled by an exponential distribution with the average amount of time equal to 17 days. Find the probability that a traveler will purchase a ticket fewer than seven days in advance. How many days do half of all travelers wait?

Solution:

Before we begin, let us make some conjectures. Why is doing so worth our time and effort? We know each of us has an idea of how the world "works." If we express this belief, and if it should turn out our conjecture is incorrect, it suggests we may not fully understand a situation as well as we think we do, and that might indicate more factors are at play than we had initially believed.

It seems the probability a traveler would wait till seven days prior to departure to purchase a ticket would be fairly low; perhaps 0.2 might be a good conjecture. The reasons for this are manifold; among them is the widespread belief that buying a ticket "at the last minute" is more expensive than purchasing well in advance.

Based upon the wording of the problem, it would appear (if our understanding of the notion of "average" is correct) half the travelers would purchase their tickets by the time we are within 17 days of departure.

Let us now put these conjectures to the test and then analyze our results.

Recall the probability density function for the Exponential Distribution is:

$$f(x) = \lambda * e^{(-\lambda * x)}, \text{ for } x \geq 0$$

where λ is the rate parameter, which represents the average number of events per unit of time, and $\lambda = \frac{1}{\mu}$. Here, $\mu = 17$, so $\lambda = \frac{1}{17}$, and our formula becomes $f(x) = \frac{1}{17} e^{-\frac{x}{17}}$.

The probability an individual would purchase their ticket seven or fewer days prior to their trip is found using the cumulative density function,

$$P(x < 7) = 1 - e^{(-7/17)} \approx 1 - 0.6625 = 0.3375$$

There is a 33.75% probability an individual will purchase their tickets within seven days of departure. This is higher than our prediction, but not unreasonably so; we simply misjudged how willing people were to delay their ticket purchase.

Now we must compute the number of days that half of the travelers will wait, which is the same as the data's 50th percentile. That is, we must solve:

$$0.50 = 1 - e^{-\frac{x}{17}}$$

$$e^{-\frac{x}{17}} = 0.50$$

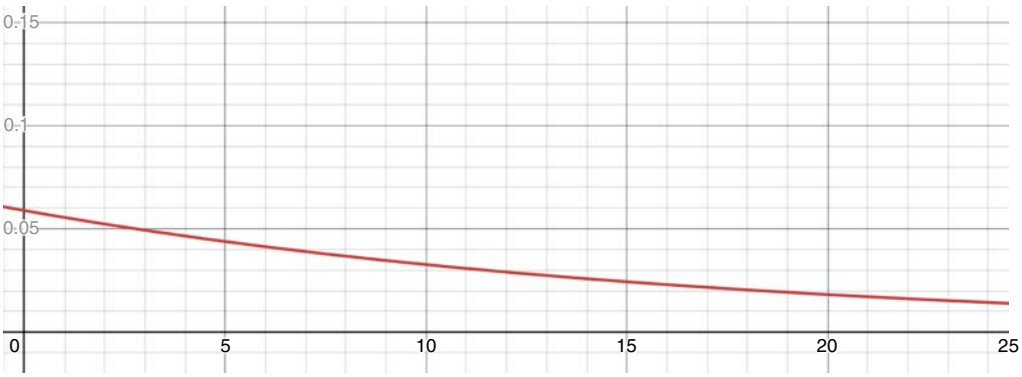
$$-\frac{x}{17} = \ln 0.50$$

$$-\frac{x}{17} \approx -0.693$$

$$x \approx 11.78$$

This is interesting; though the average number of days ahead of departure a person will purchase plane tickets is 17 days, 50% of travelers will purchase their tickets within 11.78 days of departure. What would cause this to be the case?

Remember, the data is following an exponential distribution, which would look like this:



The tail of the graph to the right is decaying exponentially; this means most of the data is at the beginning of the span of time (in days), and so 50% of the area under the curve is likely to the left of 11.78 days.

11.8.6 Expected Value and Variance of Continuous Random Variables

11.8.6.1 Expected Value of Continuous Random Variables

As was the case for discrete random variables, the expected value, or mean, of a continuous random variable X is denoted as $E(X)$ or μ and represents the average or central value of the distribution.

To calculate the expected value, we can divide the range of the random variable into multiple subintervals and calculate the weighted average of the midpoints of each subinterval.

Example 11.47 Suppose the weight of adult individuals in a certain population is modeled by a Normal Distribution with $\mu = 70$ kg and $\sigma = 5$ kg.

Solution:
To calculate the expected value of the weight, we can divide the range of weights into multiple subintervals and use the following steps:

Divide the range of weights into subintervals, for example, [55, 60], [60, 65], [65, 70], [70, 75], [75, 80], and [80, 85].

Find the midpoint of each subinterval, for example, 57.5 kg, 62.5 kg, 67.5 kg, 72.5 kg, 77.5 kg, and 82.5 kg.

Estimate the average height of the probability density function (f(x)) for each subinterval, for example, 0.0020, 0.0112, 0.0400, 0.0484, 0.0208, and 0.0038.

Calculate the area of each subinterval, for example, 0.0100, 0.0560, 0.2000, 0.2420, 0.1040, and 0.0190.

Multiply the midpoint of each subinterval by its corresponding area to get the contribution to the expected value.

Sum the contributions and divide by the total area under the curve to get the approximate expected value.

Performing these steps, we get an approximate expected value of 70.00 kg, which is the same as the given parameter μ.

As we saw earlier, appealing to the results of calculus (which takes the process of subdivision of the range of weights into multiple subintervals and allows the number of subintervals to become infinite) allows us to devise a formula for the expected value of a continuous random variable. Unfortunately, doing so here is beyond the scope of our conversation and is best left to a course specifically addressing advanced statistics and incorporating integral calculus.

11.8.6.2 Variance of Continuous Random Variables

The variance, denoted as Var(X) or σ^2, is a measure of the dispersion or spread of a continuous random variable X. To calculate the variance, we can divide the range of the random variable into multiple subintervals and calculate the weighted average of the squared deviations from the mean.

Example 11.48 Continuing the previous example, let us calculate the variance of the weight of adult individuals in the population.

Solution:
We can divide the range of weights into multiple subintervals and use the following steps:

Divide the range of weights into subintervals, for example, [55, 60], [60, 65], [65, 70], [70, 75], [75, 80], and [80, 85].

Find the midpoint of each subinterval, for example, 57.5 kg, 62.5 kg, 67.5 kg, 72.5 kg, 77.5 kg, and 82.5 kg.

Calculate the squared deviation of each midpoint from the mean (70 kg), for example, 144.75, 81.25, 25.00, 6.25, 81.25, and 144.75.

Estimate the average height of the probability density function (f(x)) for each subinterval, for example, 0.0020, 0.0112, 0.0400, 0.0484, 0.0208, and 0.0038.

Calculate the area of each subinterval, for example, 0.0100, 0.0560, 0.2000, 0.2420, 0.1040, and 0.0190.

Multiply the squared deviation of each midpoint by its corresponding area to get the contribution to the variance.

Sum the contributions and divide by the total area under the curve to get the approximate variance.

Performing these steps, we get an approximate variance of 25.00 kg^2, which is the same as the given parameter σ^2. Recalling the standard deviation is the square root of the variance, we realize immediately the standard deviation in this case is 5 kg.

11.9 The Central Limit Theorem

The Central Limit Theorem (CLT) is a fundamental concept in probability theory and statistics that describes the behavior of the sum or average of many independent and identically distributed random variables. It is a powerful tool that allows us to make inferences about the mean and distribution of a population based on a sample drawn from that population.

11.9.1 Understanding the Central Limit Theorem

The CLT states that as the sample size (n) increases, the distribution of the sample mean ($\overline{x}$) will approach a normal distribution, regardless of the underlying distribution of the individual random variables. This holds true even if the original random variables do not follow a normal distribution.

That is, the CLT is a statistical theory that states that if you take a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from that population will be roughly equal to the population mean.

All that has been said here is correct, but we require a more intuitive sense of what is happening. To this point, we have been given data and asked to calculate descriptive statistics about that data.

Now, the question is being reversed.

Typically, the situation with which we are confronted is that we do not know the values of the parameters of the population: the population mean and the population standard deviation. We can obtain the related statistics for samples, and the issue is if, based upon those statistics, we can project with any level of confidence the values of the parameters.

This is the focus of the CLT.

The key assumptions for the CLT are:

- The random variables are independent and identically distributed (i.i.d.).
- The random variables have a finite mean (μ) and finite variance (σ^2).
- The sample size (n) is sufficiently large (typically $n \geq 30$) and is the same for each of the samples.

Under these conditions, the distribution of the sample mean ($\overline{x}$) will approach a normal distribution with a mean equal to the population mean (μ) and a variance equal to the population variance ($\sigma 2$) divided by the sample size (n). In other words, if you took many samples of size n and calculated the mean and standard deviation of each of those, then the collection of all those parameters would form a normal distribution.

Mathematically, the CLT can be expressed as:

As $n \to \infty$, the distribution of $(\overline{x} - \mu)/(\sigma/\sqrt{n})$ approaches a standard normal distribution (N(0, 1)).

This means that as the sample size increases, the standardized sample mean, $(\overline{x} - \mu)/(\sigma/\sqrt{n})$, follows a standard normal distribution, N(0, 1). Observe there is no specification of the type of distribution the data itself possesses.

11.9.2 Visualizing the Central Limit Theorem

To illustrate the CLT, let's consider an example using a continuous random variable with a non-normal distribution.

Example 11.49 Suppose the weight of adult individuals in a certain population is modeled by an Exponential Distribution with a parameter $\lambda = 0.1$. We want to investigate the behavior of the sample mean as the sample size increases.

Small Sample Size ($n = 5$):

- Draw five random samples from the Exponential Distribution with $\lambda = 0.1$.
- Calculate the sample mean ($\overline{x}$) for these five samples.
- Repeat this process many times (e.g., 1000 times) to obtain the distribution of the sample mean.

The distribution of the sample mean with a small sample size ($n = 5$) is likely to be skewed, as the Exponential Distribution is a skewed distribution.

Larger Sample Size ($n = 30$):

1) Draw 30 random samples from the Exponential Distribution with $\lambda = 0.1$.
2) Calculate the sample mean ($\overline{x}$) for these 30 samples.
3) Repeat this process many times (e.g., 1000 times) to obtain the distribution of the sample mean.

The distribution of the sample mean with a larger sample size ($n = 30$) will begin to approach a normal distribution, as predicted by the Central Limit Theorem.

Even Larger Sample Size ($n = 100$):

1) Draw 100 random samples from the Exponential Distribution with $\lambda = 0.1$.
2) Calculate the sample mean ($\overline{x}$) for these 100 samples.
3) Repeat this process many times (e.g., 1000 times) to obtain the distribution of the sample mean.

The distribution of the sample mean with an even larger sample size ($n = 100$) will be very close to a normal distribution, as the Central Limit Theorem becomes more pronounced.

Example 11.50 Suppose the population of 20-year-old women in a particular state have weights with a distribution of an unknown type. If this population of women have a mean weight of 128 pounds with a standard deviation of 30 pounds, what can we conclude about the mean and standard deviation of samples of 144 such women?

Solution:
Recall the key assumptions for the Central Limit Theorem are:

- The random variables are independent and identically distributed (i.i.d.).
- The random variables have a finite mean (μ) and finite variance (σ^2).
- The sample size (n) is sufficiently large (typically $n \geq 30$).

The sample is random, and the observations are for different individuals, so this situation satisfies part A. The variables have a finite mean and variance, and we are told (though the distribution is not "normal") its size is 144, which qualifies as a "large sample size."

Thus, the CLT applies, from which we conclude the mean of the samples of these women will be 128 pounds, the same as that of the population.

To find the standard deviation of the sample means, we calculate

$$\frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{144}} = \frac{30}{12} = 2.5 \text{ pounds}$$

This tells us that, for samples of size 144, the mean of the samples should be (typically) within 2.5 pounds of the population standard deviation.

Typically, the problem is the reverse of that described in the above scenario.

Example 11.51 Suppose we are interested in determining the percentage of cell phone customers who exceed the time allowance on a basic cell phone contract. It is believed that (for those people who exceed the time allowance) the excess time used has a mean of 22 minutes, with a standard deviation of 20 minutes. A random sample of 90 customers who exceeded their time allowance is gathered; what is the probability we find their usage exceeded their time allowance by more than 18 minutes?

Solution:

By the Central Limit Theorem, the distribution of the average time in excess of the contracted amount will follow a normal distribution, with mean $\overline{x} = 22$ and standard deviation $s = \frac{20}{\sqrt{90}} \approx 2.108$. The value of 18 minutes has a z-score of $\frac{-4}{2.108} \approx -1.90$, which tells us it is 1.9 standard deviations below the mean. Consultation of a standard normal distribution chart tells us the probability such a sample would have been found is 0.0287.

12

Graph Theory

12.1 What Is Graph Theory?

Graph theory is a branch of discrete mathematics that studies the properties and applications of **graphs**. A graph is a mathematical structure that consists of a set of vertices (also known as nodes) and a set of edges that connect pairs of vertices. Graph theory has numerous applications in computer science, engineering, social sciences, and other fields.

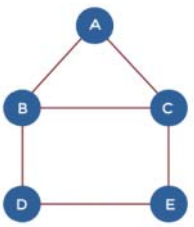
12.1.1 Definition of a Graph

A graph G is defined as an ordered pair (V, E), where:

- V is a finite, non-empty set of vertices (or nodes).
- E is a set of edges, where each edge is an unordered pair of vertices (u, v), with $u, v \in V$.

In other words, a graph is a collection of vertices, some (possibly all) of which are connected by edges. The vertices represent objects or entities, while the edges represent relationships or connections between those objects or entities.

Here is an example of a graph:



The vertices in this graph are labeled as A through E, and the edges are represented by the line segments connecting the vertices. As an illustration of potential relationships illustrated here, suppose A = Alphonse, B = Betty, C = Cerano, D = Dao, E = Elise, and the edges represent friendship relationships. Here, you see Betty is friends with Dao, Alphonse, and Cerano, but not with Elise.

There are two main types of graphs:

1) Undirected graphs: In an undirected graph, the edges are bidirectional, meaning that if there is an edge (u, v), then there is also an edge (v, u). The edges have no specific direction or orientation. The graph in the first example given above is an undirected graph.

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

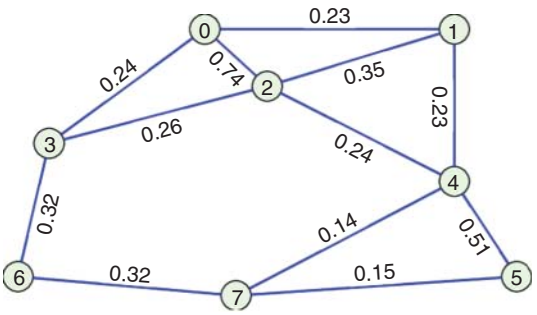
2) Directed graphs (digraphs): In a directed graph, the edges have a specific direction or orientation. If there is an edge (u, v), it means that there is a directed connection from vertex u to vertex v, but not necessarily from v to u.

Example 12.1 Suppose we have a graph representing a social media network, where the vertices represent users, and an edge (u, v) indicates that user u follows user v. This is a directed graph because the "follows" relationship is not necessarily symmetric. User u may follow user v, but user v may not follow user u.

Graphs can also have additional properties, such as:

- Weighted graphs: In a weighted graph, each edge is associated with a numerical value called a weight, which represents some property or cost of the connection.
- Labeled graphs: In a labeled graph, each vertex or edge is associated with a label or attribute that provides additional information about the object or relationship.

Example 12.2 Consider a graph representing a road network, where the vertices represent cities, and the edges represent roads connecting the cities. The weights on the edges could represent the distances between the cities. Since the distance is unchanged if the direction traveled between the cities is reversed, this is an example of a weighted, undirected graph.



The graph shown represents the distance between the homes of various individuals (along paved roads), measured in miles. The individuals are labeled with numbers from 0 to 7, inclusive. Note there is no direct edge connecting vertex 7 to vertex 2; this is not to say there is no way to go from vertex 7 to vertex 2; rather, there are at least two distinct paths, each of which passes the home of an "intermediate" individual.

As an illustration, one path from vertex 7 to vertex 2 traverses a route that passes by vertex 4 along the way and is a distance of 0.36 miles; a second path traverses a route that passes vertices 6 and 3 along the way and has a distance of 0.90 miles. There are other paths between these two vertices.

Graph theory provides a rich set of concepts, algorithms, and theorems to analyze and manipulate graphs, such as:

- Connectivity: Determining whether a graph is connected (i.e., there is a path between every pair of vertices) or disconnected.
- Paths and cycles: Studying the existence and properties of paths (sequences of vertices connected by edges) and cycles (paths that start and end at the same vertex).

- Shortest paths: Finding the shortest path between two vertices in a weighted graph, using algorithms like Dijkstra's algorithm or the Bellman–Ford algorithm.
- Graph coloring: Assigning colors to vertices such that no two adjacent vertices have the same color, which has applications in scheduling and optimization problems.

These are just a few examples of the concepts and problems studied in graph theory. The field is vast and has numerous practical applications in various domains, making it an essential part of discrete mathematics.

12.1.2 Brief History of Graph Theory

Graph theory has a rich and fascinating history that dates back to the 18th century. The development of graph theory has been driven by various real-world problems and has contributed to numerous fields, including mathematics, computer science, and engineering. Here, we will explore some of the key moments and figures in the history of graph theory.

12.1.2.1 Leonhard Euler and the Königsberg Bridge Problem (1736)

Graph theory is often said to have begun with the work of the Swiss mathematician Leonhard Euler. In 1736, Euler solved the famous Königsberg Bridge Problem, which asked whether it was possible to walk through the city of Königsberg (now Kaliningrad, Russia) by crossing each of its seven bridges exactly once and returning to the starting point.

Euler proved that such a walk was impossible by representing the problem as a graph, with land masses as vertices and bridges as edges. He showed that for a walk of this kind to exist, each vertex must have an even number of edges incident to it, except for the starting and ending vertices, which may have an odd number of edges. This condition is now known as Euler's Theorem.

12.1.2.2 Gustave Kirchhoff and the Study of Trees (1845)

German physicist Gustav Kirchhoff introduced the concept of a tree in graph theory while studying electrical networks. A tree is a connected graph with no cycles. Kirchhoff used trees to analyze electrical circuits and developed the fundamental laws of circuit analysis, known as Kirchhoff's laws.

12.1.2.3 Arthur Cayley and the Enumeration of Trees (1857)

British mathematician Arthur Cayley studied the problem of counting the number of trees with a given number of vertices. He derived a formula for the number of labeled trees with n vertices, known as Cayley's formula. This work laid the foundation for the field of enumerative combinatorics.

12.1.2.4 William Hamilton and the Icosian Game (1857)

Irish mathematician William Rowan Hamilton invented the Icosian Game, which involves finding a cycle that visits each vertex of a dodecahedron exactly once. This type of cycle is now known as a Hamiltonian cycle, and the problem of determining whether a graph contains a Hamiltonian cycle is a fundamental problem in graph theory.

12.1.2.5 Dénes Kőnig and the First Graph Theory Book (1936)

Hungarian mathematician Dénes Kőnig published the first book entirely devoted to graph theory, titled "Theorie der endlichen und unendlichen Graphen" (Theory of Finite and Infinite Graphs). This book systematically developed many of the fundamental concepts and results in graph theory.

12.1.2.6 Claude Shannon and the Application of Graph Theory to Communication Networks (1948)

American mathematician and engineer Claude Shannon used graph theory to study communication networks. He introduced the concept of information entropy and laid the foundations for information theory. Shannon's work demonstrated the practical applications of graph theory in the design and analysis of communication systems.

12.1.2.7 Paul Erdős and the Rise of Modern Graph Theory (1950s–1990s)

Hungarian mathematician Paul Erdős made significant contributions to graph theory and combinatorics. He introduced many important concepts, such as the Erdős–Rényi random graph model and the Ramsey theory. Erdős collaborated with numerous mathematicians and helped to establish graph theory as a major area of research in the second half of the 20th century.

These are just a few highlights from the history of graph theory. The field has continued to grow and evolve, with new applications and challenges emerging in various domains, such as computer networks, social networks, and bioinformatics. The study of graph theory remains an active and vital area of research in discrete mathematics and related fields.

12.1.3 Applications of Graph Theory

12.1.3.1 Computer Science

In computer science, graph theory is used extensively in the design and analysis of algorithms, data structures, and computer networks.

Example 12.3 Computer Networks

Computer networks can be modeled as graphs, where the vertices represent computing devices (such as computers, routers, or switches), and the edges represent the connections between them (such as Ethernet cables or wireless links). Graph theory can be used to study the properties of these networks, such as connectivity, data flow, and network security.

Example 12.4 Shortest Path Algorithms

Graph theory is used in the development of shortest path algorithms, such as Dijkstra's algorithm and the Bellman–Ford algorithm. These algorithms find the shortest path between two vertices in a weighted graph, which has applications in navigation systems, routing protocols, and optimization problems.

12.1.3.2 Social Networks

Graph theory is extensively used in the analysis of social networks, where vertices represent individuals or entities, and edges represent the relationships or interactions between them.

Example 12.5 Friendship Networks

A friendship network can be modeled as a graph, where each vertex represents a person, and an edge between two vertices indicates that the corresponding individuals are friends. Graph theory can be used to study the properties of these networks, such as the degree of separation between individuals (i.e., the "six degrees of separation" concept), the identification of communities or clusters, and the spread of information or influence through the network.

12.1.3.3 Biology

Graph theory finds application in biology, particularly in the study of biological networks, such as protein–protein interaction networks, metabolic networks, and gene regulatory networks.

Example 12.6 Protein–Protein Interaction Networks

Proteins in a cell interact with each other to perform various biological functions. These interactions can be represented as a graph, where the vertices represent proteins, and the edges represent the interactions between them. Graph theory can be used to identify important proteins (such as hub proteins that have a high degree of connectivity), find functional modules or pathways, and study the robustness of the network to perturbations.

12.1.3.4 Chemistry

In chemistry, graph theory is used to represent and analyze molecular structures, chemical reactions, and crystal structures.

Example 12.7 Molecular Graphs

A molecule can be represented as a graph, where the vertices represent atoms, and the edges represent chemical bonds. Graph theory can be used to study the properties of these molecular graphs, such as the connectivity, symmetry, and stability of the molecules. This has applications in drug design, materials science, and computational chemistry.

12.1.3.5 Other Fields

Graph theory has applications in numerous other fields, including:

- Physics: Graph theory is used in the study of phase transitions, percolation theory, and network dynamics.
- Linguistics: Graphs can be used to represent the relationships between words, sentences, or concepts in natural language processing and computational linguistics.
- Transportation: Graph theory is used in the design and optimization of transportation networks, such as road networks, airline routes, and public transportation systems.
- Electrical Engineering: Graphs are used to represent and analyze electrical circuits, power grids, and communication networks.
- Operations Research: Graph theory is used in the formulation and solution of optimization problems, such as the traveling salesman problem, facility location problems, and resource allocation problems.

As research into graph theory continues to evolve, new applications are likely to emerge in various domains, demonstrating the expanding power and versatility of graph-theoretic concepts and methods.

12.2 Exploration of Basic Concepts

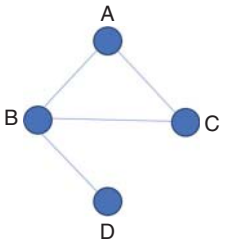
12.2.1 Vertices (Nodes)

Properties of vertices:

Degree: The degree of a vertex is the number of edges incident to it. In an undirected graph, the degree of a vertex is the number of edges connected to it. In a directed graph, we distinguish

between the in-degree (the number of incoming edges) and the out-degree (the number of outgoing edges).

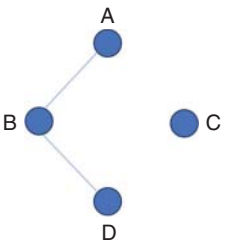
Example 12.8 In the graph below, vertex A has a degree of 3, vertex B has a degree of 2, and vertices C and D have a degree of 2.



Adjacent vertices: Two vertices are said to be adjacent if they are connected by an edge. In other words, if there is an edge (u, v) in the graph, then vertices u and v are adjacent.

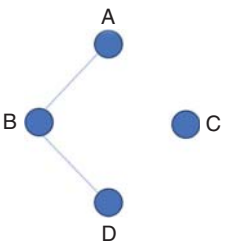
Example 12.9 In the graph above, vertex A is adjacent to vertices B and C, but not to vertex D.
Isolated vertex: An isolated vertex is a vertex with a degree of 0, meaning it has no edges incident to it.

Example 12.10 In the graph below, vertex C is an isolated vertex.



Pendant vertex: A pendant vertex, also known as a leaf vertex, is a vertex with a degree of 1, meaning it has only one edge incident to it.

Example 12.11 In the graph below, vertices A and D are pendant vertices.



Vertex labels or weights: In some graphs, vertices may have associated labels or weights that provide additional information about the objects they represent. These labels or weights could represent attributes, costs, or any other relevant data.

Example 12.12 In a graph representing a road network, the vertices could be labeled with the names of the cities they represent, and the weights could represent the population of each city.

Vertices are the basic elements of a graph, and their properties and relationships define the structure and characteristics of the graph. By studying the properties of vertices and their connections, we gain insights into the behavior and characteristics of the systems or networks represented by the graph.

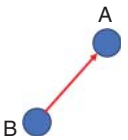
12.2.2 Edges

12.2.2.1 Directed Edges

As mentioned, in a directed graph (also known as a digraph), edges have a specific direction or orientation, indicating a one-way relationship from one vertex to another. A directed edge (u, v) indicates a connection from vertex u to vertex v, but not necessarily from v to u. Visually, the direction is indicated with an arrowhead showing the direction of the edge.

Example 12.13 Consider a graph representing a social media network, where vertices represent users, and edges represent the "follows" relationship. A directed edge (Alice, Bob) indicates that Alice follows Bob, but it does not imply that Bob follows Alice. On the other hand, a directed edge (Bob, Alice) indicates Bob follows Alice, but not vice versa.

Bob (B) ---> Alice (A)



Properties of directed edges:

Head and Tail: In a directed edge (u, v), vertex u is called the tail (or source), and vertex v is called the head (or target).

In-degree and Out-degree: The in-degree of a vertex is the number of incoming edges, while the out-degree is the number of outgoing edges. In the example above, Bob has an out-degree of 1 and an in-degree of 0, while Alice has an out-degree of 0 and an in-degree of 1.

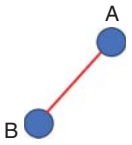
Directed Path: A directed path is a sequence of vertices connected by directed edges, where the direction of the edges follows the order of the sequence. In the example above, there is a directed path from Bob to Alice, but not from Alice to Bob.

It is an interesting feature of directed edges in relationships that we dance upon nuances of language. If, for instance, the relationship edge shown in the above example is "admires," then we are told Bob admires Alice, but not that Alice admires Bob. Alice may, in fact, detest Bob. We must take great care in interpreting the relationships indicated by directed graphs.

12.2.2.2 Undirected Edges

In an undirected graph, edges have no specific direction or orientation. An undirected edge (u, v) represents a symmetric or bidirectional relationship between vertices u and v. If there is an edge (u, v), then there is also an edge (v, u).

Example 12.14 Consider a graph representing friendships between people, where vertices represent individuals, and edges represent friendships. An undirected edge (Alice, Bob) indicates that Alice and Bob are friends, and the friendship is mutual.



Properties of undirected edges:

1) Symmetric Relationship: If there is an undirected edge (u, v), then there is also an edge (v, u). The relationship represented by an undirected edge is symmetric.
2) Degree: In an undirected graph, the degree of a vertex is the number of edges incident to it. In the example above, both Alice and Bob have a degree of 1.
3) Path: A path in an undirected graph is a sequence of vertices connected by edges, without considering the direction of the edges. In the example above, there is a path from Alice to Bob and from Bob to Alice.

Both directed and undirected edges can have associated weights or costs, representing some quantitative aspect of the relationship between the vertices. These weights could represent distance, time, capacity, or any other relevant measure.

Example 12.15 In a graph representing a road network, the undirected edges could have weights indicating the distance between the cities connected by the roads.



In this example, the edge (A, B) has a weight of 4, the edge (A, C) has a weight of 3, and the edge (B, C) has a weight of 2.

Edges, whether directed or undirected, are fundamental components of graphs and play a crucial role in defining the structure and properties of the graph. By analyzing the edges and their properties, we can gain insights into the relationships and interactions between the vertices in the graph.

Example 12.16 Navigation on the internet is an illustration of vertices and directed edges. Each website on the internet can be viewed as representing a web page. A hyperlink on a website creates links, or connections, between pages on the internet. A hyperlink can be a piece of text, an image, an icon, or a graphic that, when you click on it, points to and navigates you to a different webpage or document. It can also point to a specific section or element within the same webpage or document.

Directed edges between vertices represent hyperlinks. Note that, while it is possible to use your operating system to return to a prior website, a hyperlink is a one-way travel pathway.

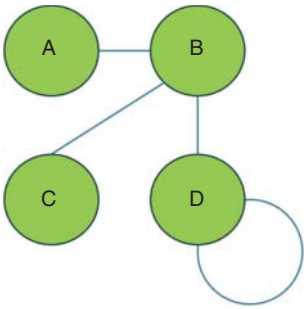
12.2.3 Loops and Multiple Edges

In graph theory, loops and are special cases of edges that can occur in graphs. The concept is important to understand as it can affect the properties and behavior of the graph and also ties closely to the notion of a programming loop.

12.2.3.1 Loops

A loop is an edge that connects a vertex to itself. In other words, it is an edge that starts and ends at the same vertex. Loops are also known as self-loops, self-edges, or degenerate edges.

Example 12.17 Consider the following graph with a loop:



In this graph, vertex D has a loop, as there is an edge that connects D to itself.

Loops can be present in both directed and undirected graphs. In a directed graph, a loop is represented by an arrow that points from a vertex back to itself.

In an undirected graph, a loop contributes twice to the degree of the vertex it is incident to. In the example above, vertex D has a degree of 3 (one from the edge connecting it to B and two from the loop). In the adjacency graph, the 1's corresponding to loops appear on the main diagonal of the matrix.

There are many instantiations of loops in applications, but let us briefly consider an appearance within a Markov Chain. Recall that a Markov Chain is a process in which the probability of an event depends upon the state of a system immediately prior to the event. In such a process, a loop would represent a steady-state element of the system.

Loops can impact the connectivity and traversal of the graph. In some cases, loops may need special handling in graph algorithms.

12.2.3.2 Adjacent Vertices

In graph theory, two vertices are said to be adjacent if they are connected by an edge. In other words, if there is an edge between two vertices, they are considered adjacent.

A graph in which every vertex is connected to every other vertex (that is, in which every vertex is adjacent to every other vertex) is called complete. The number of edges in a complete graph with n vertices is the sum of the whole numbers from 1 to $n-1$; that is, the number of edges in a complete graph having n vertices is $1+2+3+\cdots+(n-1)$.

Example 12.18 A complete graph having eight vertices must have how many edges?

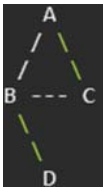
Solution:

The number of edges is the sum of the whole numbers from 1 to 7; that is, $1+2+3+4+5+6+7 = 28$.

The concept of adjacency is fundamental in understanding the relationships and connections within a graph. An example of this can be seen if we view the nodes as representing the tasks, and there is an edge $j \to k$ if and only if task j must be completed before task k. This establishes a hierarchy of steps needed to complete an algorithm.

As an illustration, we can use a directed graph to represent what must be done to complete a college degree within a particular major field of study, with the nodes representing the courses to be taken and the edges representing prerequisites.

Example 12.19 Consider the following undirected graph:



In this graph:

- Vertex A is adjacent to vertices B and C.
- Vertex B is adjacent to vertices A, C, and D.
- Vertex C is adjacent to vertices A and B.
- Vertex D is adjacent only to vertex B.

Note this is not a connected graph; A is not adjacent to D nor is C adjacent to D. Either of these situations would be sufficient to prevent the graph from being connected.

In the context of directed graphs, the notion of adjacency can be further refined based on the direction of the edges.

If there is a directed edge from vertex u to vertex v, then u is said to be adjacent to v, and v is said to be adjacent from u.

Example 12.20 Consider the following directed graph:



In this graph:

- Vertex A is adjacent to vertex B.
- Vertex B is adjacent to vertex C.
- Vertex C is adjacent to vertex D.
- Vertex D is adjacent to vertex A.
- Vertex B is adjacent from vertex A.
- Vertex C is adjacent from vertex B.
- Vertex D is adjacent from vertex C.
- Vertex A is adjacent from vertex D.

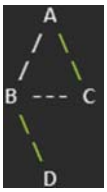
The adjacency relationship between vertices can be represented using various data structures, such as an adjacency matrix or an adjacency list.

12.2.3.3 Adjacency Matrix

An adjacency matrix is a square matrix A of size $n \times n$, where n is the number of vertices in the graph. The entry A[i][j] represents the presence or absence of an edge between vertices i and j; we must differentiate between the cases of undirected and directed graphs, however, as a distinction between the definition of the matrix entries exists.

In an undirected graph, $A[i][j] = 1$ if there is an edge between vertices i and j, and $A[i][j] = 0$ otherwise. The adjacency matrix is **symmetric**, that is, $A[i][j] = A[j][i]$. If you have not worked with matrices in any detail, an example will make this definition evident.

Example 12.21 Produce the adjacency matrix for the undirected graph



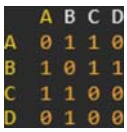
Recall we had found the following adjacency relationships for this graph:

- Vertex A is adjacent to vertices B and C.
- Vertex B is adjacent to vertices A, C, and D.
- Vertex C is adjacent to vertices A and B.
- Vertex D is adjacent only to vertex B.

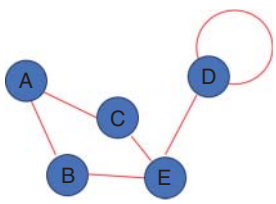
According to the definition, if we were to lay out the square matrices so the rows across are labeled with the vertices (top to bottom) and the columns with the vertices in the same order (left to right), we would place a "1" wherever an adjacency relationship exists.

That is, since A is adjacent to both B and C, there will be a "1" positioned in the A, B location, the A, C location, and (by symmetry) in the B, A and C, A locations. All other entries will be zeros.

Therefore, the adjacency matrix would be:



Example 12.22 Produce the adjacency matrix for the following graph:



Solution:

We first establish the adjacency relationships:

1) A is adjacent to B and C.
2) B is adjacent to A and E.
3) C is adjacent to A and E.
4) D is adjacent to D and E.
5) E is adjacent to B, C, and D.

We now construct a 5×5 matrix (since there are five vertices), labeling the rows and columns as indicated in the previous example.

	A	B	C	D	E
A					
B					
C					
D					
E					

1's are placed in all entries where an adjacency relationship exists:

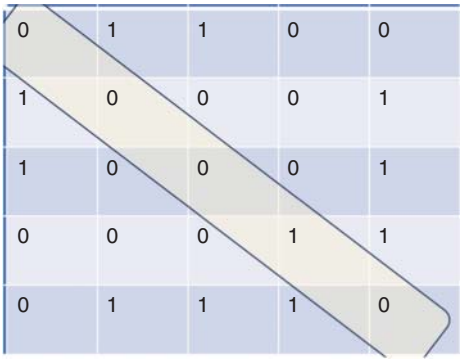
	A	B	C	D	E
A		1	1		
B	1				1
C	1				1
D				1	1
E		1	1	1	

0's are placed in all other entries:

	A	B	C	D	E
A	0	1	1	0	0
B	1	0	0	0	1
C	1	0	0	0	1
D	0	0	0	1	1
E	0	1	1	1	0

Omitting the labeling columns, the entries in the light blue area are the adjacency matrix.

We have mentioned the adjacency matrix is symmetric; let us elaborate upon this. In a square matrix, the entries lying on the diagonal from the upper left to lower right corners form what we call the **main diagonal** of the matrix.



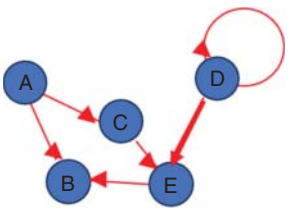
Observe the entries on either side of the matrix are mirror images of one another. This is what we mean by a symmetric matrix: entry $A_{i,j}$ = entry $A_{j,i}$.

Let us now distinguish the case for directed graphs.

In a directed graph, $A[i][j] = 1$ **if there is a directed edge from vertex *i* to vertex *j***, and $A[i][j] = 0$ otherwise. The adjacency matrix may not be symmetric in this case.

Consider the following modification to the previous example.

Example 12.23 Find the adjacency matrix for the directed graph:



There are directed edges from A to B, from A to C, from C to E, from D to D, from D to E, and from E to B. We place a "1" in the entry in row A, column B, for instance, because there is a directed edge from A to B. Zeros are placed elsewhere. Thus, we obtain:

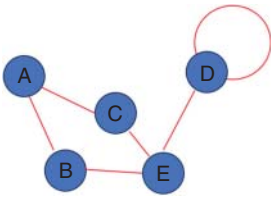
	A	B	C	D	E
A	0	1	1	0	0
B	0	0	0	0	0
C	0	0	0	0	1
D	0	0	0	1	1
E	0	1	0	0	0

Observe the adjacency matrix for the directed graph is NOT symmetric. Moreover, we see evidence here of a fact we mentioned earlier in a parenthetical remark: there is a "1" on the main diagonal, in the entry aligning with row D and column D. This is the "loop" we see in the graph, where D is connected to itself by an edge.

A bit of tedium has existed in the clumsy manner in which we have been describing adjacency properties. This is made more compact through the introduction of the concept of an Adjacency List.

An adjacency list is a collection of sublists, where each sublist corresponds to a vertex in the graph. The list for a vertex contains all the vertices that are adjacent to it.

Example 12.24 For the undirected graph shown:



the adjacency list would be:

A: [B, C]
B: [A, E]
C: [A, E]
D: [D, E]
E: [B, C, D]

Though not a huge leap forward, it does eliminate the need for individual references such as "A is adjacent to B and C."

12.2.3.4 Incident Edges

An edge is said to be **incident to a vertex** if the vertex is one of the endpoints of the edge. In other words, if an edge connects two vertices, it is incident to both of those vertices. The concept of incidence is closely related to the concept of adjacency, as incident edges determine which vertices are adjacent to each other.

Example 12.25 Consider the following undirected graph:



In this graph:

The edges (A, B) and (A, C) are incident to vertex A.
The edges (A, B), (B, C), and (B, D) are incident to vertex B.
The edges (A, C) and (B, C) are incident to vertex C.
The edge (B, D) is incident to vertex D.

In the context of directed graphs, the notion of incidence can be further refined based on the direction of the edges:

If there is a directed edge from vertex u to vertex v, then the edge is said to be incident from vertex u and incident to vertex v.

Example 12.26 Consider the following directed graph:



In this graph:

The edge (A, B) is incident from vertex A and incident to vertex B.
The edge (B, C) is incident from vertex B and incident to vertex C.
The edge (C, D) is incident from vertex C and incident to vertex D.
The edge (D, A) is incident from vertex D and incident to vertex A.

The incidence relationship between edges and vertices can be represented using an incidence matrix.

12.2.3.5 Incidence Matrix

An incidence matrix is a matrix I of size $n \times m$, where n is the number of vertices, and m is the number of edges in the graph. The entry I[i][j] represents the relationship between vertex i and edge j.

In an undirected graph, I[i][j] = 1 if edge j is incident to vertex i, and $I[i][j] = 0$ otherwise.

In a directed graph, $I[i][j] = 1$ if edge j is incident from vertex i, $I[i][j] = -1$ if edge j is incident to vertex i, and $I[i][j] = 0$ otherwise.

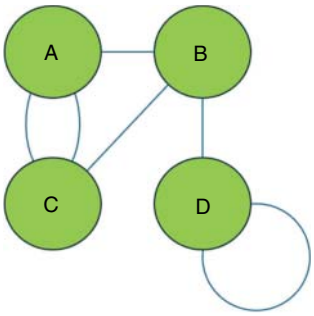
The concept of incidence is fundamental in graph theory and has various applications:

1) Degree of a Vertex: The degree of a vertex can be determined by counting the number of edges incident to it. In an undirected graph, the degree of a vertex is equal to the sum of the entries in the corresponding row of the incidence matrix.
2) Edge-Vertex Duality: The incidence relationship between edges and vertices gives rise to the concept of edge-vertex duality. In some graph problems, it may be beneficial to switch the roles of edges and vertices, treating edges as vertices and vertices as edges.
3) Graph Algorithms: The incidence information is used in various graph algorithms, such as finding Eulerian circuits or paths, determining connectivity, and more.

12.2.3.6 Multiple Edges

Multiple edges, also known as parallel edges, occur when there is more than one edge between the same pair of vertices. In other words, if there are two or more edges connecting the same pair of vertices, they are called multiple edges. The nomenclature "parallel" does not refer to edges that are visually displayed as parallel lines.

For instance, consider the graph:



There are multiple edges connecting vertex A to vertex C; those edges are said to be parallel, but there is no association with the geometric term "parallel," since the edges are drawn as arcs.

Example 12.27 Consider the following graph having multiple edges:



Why, visually, can you see the graph has multiple edges?

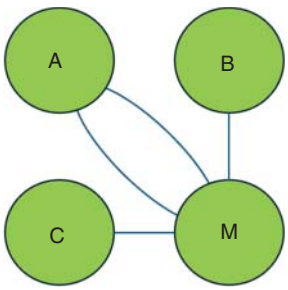
Solution:
In this graph, there are two edges connecting vertices A and B. This, by itself, establishes the existence of multiple edges, but additionally, there are two edges connecting vertices A and C, which would also establish this to be a graph with multiple edges.

Multiple edges can be present in both directed and undirected graphs. In a directed graph, multiple edges can have the same or opposite directions.

Let us explore an application of a graph having multiple edges, and see if we can recognize how it would "fit" a real-world scenario.

Imagine a computer network within which users can access the mainframe via the use of a password at their own computer terminal. In this situation, there are three users, one of whom serves in a managerial position. The individuals are identified using A, B, and C, where A is the individual acting as a manager.

We will represent the system mainframe with the identification M, with edges of the graph representing access to that mainframe.

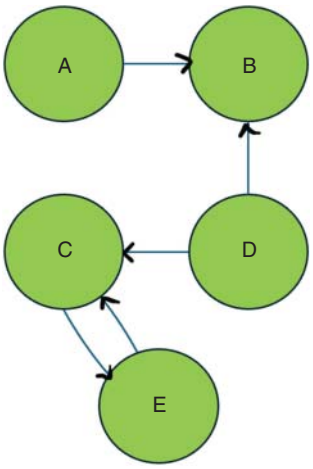


Observe that the manager has two edges connecting to the mainframe location; why would this be the case?

In the event of some sort of system failure, the manager may be granted a backdoor access path to the mainframe. In this way, were the system to somehow be corrupted in a manner that precluded standard access routes, the manager would have a secondary path of access to the mainframe, ensuring not all employees would be "locked out."

What would be an illustration of a "directed graph" having multiple edges?

Imagine a network of "friends" who are involved in office politics in such a way that conflict has arisen.



The employees are Alfred, Brian, Carlos, Dmitri, and Elena. The directional arrows on the vertices represent who is talking to whom. That is, Alfred is talking to Brian, but not vice versa. Similarly, Dmitri is speaking to Brian, but not vice versa; continuing with this interpretation, we see that only Carlos and Elena are speaking to one another.

Imagine Elena wishes to communicate with Dmitri. Is this possible? Why or why not?

Inspection of the graph reveals Elena can communicate with Carlos, but that Carlos is unwilling to speak directly to Dmitri. Carlos CAN communicate with Alfred, who can then relay the message to Brian, but Brian has no conduit via which he can communicate with Dmitri. Therefore, Elena cannot communicate with Dmitri due to the volatile office environment.

On the other hand, should Dmitri wish to communicate with Elena, he is able to contact Carlos, who is (in turn) in communication with Elena. This means there is a one-way path of communication only between Elena and Dmitri.

Can Alfred communicate with Elena?

Is it feasible, in the long run, for an organization to run under these relationship conditions? Why or why not?

Multiple edges can create ambiguity in certain graph problems, such as finding the shortest path between two vertices. In such cases, it may be necessary to specify which edge is being considered.

12.2.3.7 Handling Loops and Multiple Edges

When working with graphs that contain loops or multiple edges, it is important to consider how they should be treated based on the specific problem or algorithm being addressed. Some common approaches include:

Removing Loops In some cases or programs, loops may be removed from the graph before applying certain algorithms or solving specific problems. This simplifies the graph structure and can make the problem more manageable.

In the programming language for Mathematica, as an illustration, the "IGSimpleGraph" command collapses loops from a graph. Each language may have native commands with which users must become familiar, in order to simplify those graphs.

Collapsing Multiple Edges Multiple edges between the same pair of vertices can be collapsed into a single edge, often by aggregating the weights or attributes of the multiple edges. This simplification can make the graph easier to work with while still preserving the essential connectivity information.

In the programming language Python, the command for doing so is g.simplify.

We mention these cases only as illustrations. It is impractical to offer global methods for performing a particular task, since each language exists within its own sphere, and commands evolve over time with new iterations of the software. Going forward, we will refrain from offering such examples.

Allowing Loops and Multiple Edges In some cases, loops and multiple edges may be allowed and treated as valid parts of the graph. The algorithms and problems are designed to handle these special cases explicitly.

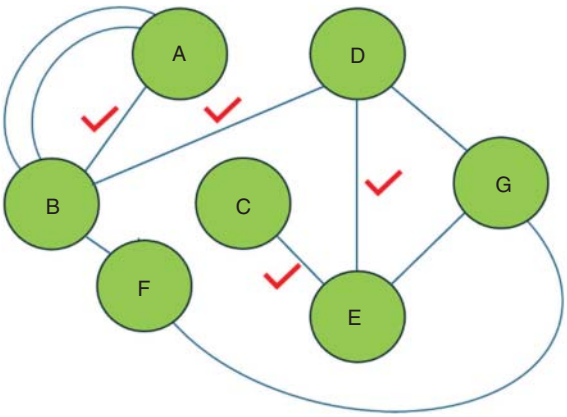
It is important to clearly specify how loops and multiple edges are handled when defining a graph problem or algorithm to ensure consistency and avoid ambiguity.

12.2.3.8 Density of a Graph

We have spoken of the number of edges of a graph (differentiating, as needed, between undirected and directed graphs). The relationship between how many edges exist between vertices and how many edges are possible between vertices is referred to as the **density** of a graph. There is, as we shall see, a related concept called the **sparsity** of a graph.

12.2.3.9 Diameter of a Graph

What is the diameter of the shown graph?



In this graph, the most distant vertices are A and C. While there are numerous paths from A to C (for example, $A-B-F-G-D-E-C$), the shortest distance traverses four edges, and this is the diameter of the graph.

Note this is not a contradiction of the set of properties given a moment ago; the graph here is NOT complete.

12.3 Types of Graphs

Graphs can be classified into different types based on their properties and characteristics. Understanding these types is essential for modeling and solving various real-world problems using graph theory. In this section, we will explore some common types of graphs, starting with undirected graphs.

12.3.1 Undirected Graphs

An undirected graph is a graph in which the edges have no specific direction or orientation. In other words, the edges are bidirectional, and the relationship between vertices is symmetric. If there is an edge between vertices u and v, it can be traversed from u to v and from v to u.

Example 12.28 Consider a graph representing friendships between people, where each person is represented by a vertex, and friendships are represented by edges. Since friendships are typically mutual, this graph would be an undirected graph.



In this graph, the edges represent the friendships between Alice, Bob, Carol, and David. The edge between Alice and Bob indicates that they are friends, and this friendship is symmetric.

Properties of Undirected Graphs:

12.4 Symmetric Adjacency: If vertex u is adjacent to vertex v, then vertex v is also adjacent to vertex u. In the adjacency matrix representation of an undirected graph, the matrix is symmetric, that is, $A[i][j]=A[j][i]$.

12.5 Edge Representation: An edge in an undirected graph is represented by an unordered pair of vertices (u, v) or (v, u). The order of the vertices does not matter, as the edge is bidirectional.

12.6 Degree: The degree of a vertex in an undirected graph is the number of edges incident to it. In the example above, Alice has a degree of 2, Bob has a degree of 2, Carol has a degree of 2, and David has a degree of 2.

12.7 Connectivity: An undirected graph is connected if there is a path between every pair of vertices. In the example above, the graph is connected, as there is a path between any two vertices.

12.8 Cycles: An undirected graph can contain cycles, which are paths that start and end at the same vertex. In the example above, the path Alice – Bob – David – Carol – Alice forms a cycle.

Undirected graphs have various applications, including:

- Social Networks: Undirected graphs can model social networks, where vertices represent individuals, and edges represent connections or relationships between them, such as friendships or collaborations.
- Computer Networks: Undirected graphs can represent computer networks, where vertices represent devices or nodes, and edges represent the communication links between them.
- Transportation Networks: Undirected graphs can model transportation networks, such as road networks or airline routes, where vertices represent locations, and edges represent the connections between them.
- Chemical Compounds: Undirected graphs can represent chemical compounds, where vertices represent atoms, and edges represent the chemical bonds between them.

When solving problems or designing algorithms for undirected graphs, it is important to consider the specific properties and characteristics of undirected graphs. For example, algorithms for finding the shortest path or determining connectivity in an undirected graph may differ from those used for directed graphs.

Undirected graphs provide a fundamental structure for modeling and analyzing relationships and connections in various domains. They offer a simple yet powerful representation that captures the symmetric nature of many real-world systems.

12.3.2 Directed Graphs (Digraphs)

A directed graph, also known as a digraph, is a graph in which the edges have a specific direction or orientation. In a directed graph, the edges are represented by ordered pairs of vertices, indicating a one-way relationship from one vertex to another. The direction of an edge is typically depicted by an arrow.

Example 12.29 Consider a graph representing a social media network, where vertices represent users, and directed edges represent the "follows" relationship. If user A follows user B, there would be a directed edge from vertex A to vertex B.



In this graph, the directed edges represent the "follows" relationships. Alice follows Bob and Carol, Bob follows Carol, Carol follows David, and David follows Bob.

Properties of Directed Graphs:

- Asymmetric Adjacency: If there is a directed edge from vertex u to vertex v, it does not necessarily imply that there is a directed edge from vertex v to vertex u. In the adjacency matrix representation of a directed graph, the matrix may not be symmetric.

- Edge Representation: A directed edge in a digraph is represented by an ordered pair of vertices (u, v), indicating an edge from vertex u to vertex v.
- In-degree and Out-degree: In a directed graph, each vertex has an in-degree and an out-degree. The in-degree of a vertex is the number of incoming edges, while the out-degree is the number of outgoing edges. In the example above, Alice has an in-degree of 0 and an out-degree of 2, Bob has an in-degree of 1 and an out-degree of 1, Carol has an in-degree of 2 and an out-degree of 1, and David has an in-degree of 1 and an out-degree of 1.
- Paths and Cycles: In a directed graph, paths and cycles follow the direction of the edges. A path in a directed graph is a sequence of vertices connected by directed edges, where the direction of the edges follows the order of the sequence. A cycle in a directed graph is a path that starts and ends at the same vertex, following the direction of the edges.
- Strongly Connected Components: A directed graph is said to be strongly connected if there is a directed path from every vertex to every other vertex. A strongly connected component (SCC) is a maximal subgraph of a directed graph in which all vertices are strongly connected.

Directed graphs have various applications, including:

- Web Graphs: Directed graphs can model web graphs, where vertices represent web pages, and directed edges represent hyperlinks between the pages.
- Social Networks: Directed graphs can represent social networks with asymmetric relationships, such as the "follows" relationship in social media platforms.
- Transportation Networks: Directed graphs can model transportation networks with one-way streets or flight routes, where the direction of the edges represents the allowed direction of travel.
- Dependency Graphs: Directed graphs can represent dependency relationships, such as task dependencies in project management or module dependencies in software development.

When solving problems or designing algorithms for directed graphs, it is important to consider the specific properties and characteristics of directed graphs. Algorithms for directed graphs may differ from those used for undirected graphs, as the direction of the edges can impact the traversal and connectivity of the graph.

Directed graphs provide a powerful tool for modeling and analyzing asymmetric relationships and one-way interactions in various domains. They allow for the representation of hierarchical structures, dependencies, and flow of information or resources in complex systems.

12.3.3 Weighted Graphs

A weighted graph is a graph in which each edge is assigned a numerical value or weight. The weight of an edge represents some quantitative attribute or cost associated with the connection between the vertices. Weighted graphs can be either undirected or directed, depending on whether the edges have a specific direction.

Properties of Weighted Graphs:

1) Edge Weights: Each edge in a weighted graph is associated with a numerical value or weight. The weight can represent various attributes such as distance, cost, time, or capacity, depending on the context of the problem.
2) Weight Representation: The weights of the edges can be represented in different ways, such as:
   - Adjacency Matrix: In the adjacency matrix representation of a weighted graph, the matrix entries A[i][j] represent the weight of the edge between vertices i and j. If there is no edge between two vertices, the corresponding matrix entry is typically set to a special value, such as 0 or infinity.

- Edge List: In the edge list representation, each edge is represented as a tuple or a list containing the source vertex, destination vertex, and the weight of the edge.

3) Shortest Paths: One of the common problems in weighted graphs is finding the shortest path between two vertices. The shortest path is the path with the minimum total weight. Various algorithms, such as Dijkstra's algorithm or the Bellman–Ford algorithm, can be used to find the shortest paths in weighted graphs.
4) Minimum Spanning Trees: A minimum spanning tree (MST) is a tree that spans all the vertices of a weighted graph with the minimum total weight of edges. MSTs are used in various applications, such as network design or clustering. Algorithms like Kruskal's algorithm or Prim's algorithm can be used to find the MST of a weighted graph.
5) Network Flow: Weighted graphs can represent flow networks, where the weights of the edges represent the capacity or maximum flow that can pass through the edges. Network flow problems, such as finding the maximum flow or minimum cut, can be solved using algorithms like the Ford–Fulkerson algorithm or the Edmonds–Karp algorithm.

Weighted graphs have various applications, including:

- Transportation Networks: Weighted graphs can model transportation networks, where the weights represent distances, travel times, or costs between locations.
- Communication Networks: Weighted graphs can represent communication networks, where the weights represent the bandwidth, latency, or signal strength of the links between nodes.
- Resource Allocation: Weighted graphs can model resource allocation problems, where the weights represent the costs or benefits associated with allocating resources to different tasks or entities.
- Optimization Problems: Weighted graphs can be used to model and solve various optimization problems, such as finding the shortest path, MST, or maximum flow in a network.

When solving problems or designing algorithms for weighted graphs, it is important to consider the specific properties and characteristics of weighted graphs. The presence of weights on the edges adds an extra dimension to the problem and may require specialized algorithms and data structures to efficiently handle the weights.

12.3.4 Simple Graphs

A simple graph is a type of undirected graph that has no loops (edges that connect a vertex to itself) and no multiple edges (more than one edge between the same pair of vertices). In other words, a simple graph is an unweighted, undirected graph without any self-loops or parallel edges.

Example 12.30 Consider a graph representing a social network of friends, where each person is represented by a vertex, and friendships are represented by edges.



In this simple graph, the vertices Alice, Bob, Carol, and David represent people, and the edges represent friendships between them. Each edge connects two distinct vertices, and there are no multiple edges or self-loops.

Properties of Simple Graphs:

- Maximum Number of Edges: In a simple graph with n vertices, the maximum number of edges is $n(n-1)/2$. This occurs when the graph is complete, meaning that every pair of vertices is connected by an edge.
- Degree: The degree of a vertex in a simple graph is the number of edges incident to it. Since there are no loops or multiple edges, the degree of a vertex is equal to the number of its adjacent vertices.

Simple graphs have various applications, including:

1) Basic Social Networks: Simple graphs can model social networks, where vertices represent individuals, and edges represent relationships or connections between them, such as friendships or collaborations.
2) Computer Networks: Simple graphs can represent computer networks, where vertices represent devices or nodes, and edges represent the communication links between them.
3) Molecular Structures: Simple graphs can model molecular structures, where vertices represent atoms, and edges represent the chemical bonds between them.
4) Geographic Maps: Simple graphs can represent geographic maps, where vertices represent locations or cities, and edges represent the roads or routes connecting them.

When solving problems or designing algorithms for simple graphs, it is important to consider the specific properties and characteristics of such graphs. The absence of loops and multiple edges simplifies certain algorithms and data structures, as there is no need to handle these special cases. They are widely used in various fields, including computer science, social network analysis, and discrete mathematics, to model and analyze systems where the focus is on the presence or absence of connections rather than their direction or multiplicity.

While simple graphs are a common and fundamental type of graph, there are situations where loops, multiple edges, or directed edges are necessary to accurately model the problem at hand. In such cases, other types of graphs, such as multigraphs or directed graphs, may be more appropriate.

12.3.5 Mixed Graphs

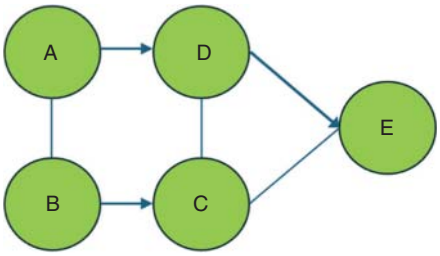
A mixed graph is a graph that can be obtained from a simple undirected graph by replacing some of the edges by edges oriented in precisely one of the two possible directions. We use G = (V(G), E(G), A(G)) to denote a mixed graph with a set of vertices V(G), a set of (undirected) edges E(G), and a set of directed edges A(G).

Mixed graphs are a relatively recent concept, first defined by N.V. Lambin and V.S. Tanaev in 1970.

As an illustration, mixed graphs represent complex social networks with high accuracy, as there may be directed edges and undirected edges, with each type indicating a particular relationship within the social network hierarchy. Directed edges are the indication of influences, dominations, or followers, while undirected edges are the indication of the opposite cases, where individuals are connected while following each other.

For example, the social media platform known as Facebook contains networks created among the users of the service. These networks allow mixed direction when, if at least one friend follows another along with their friendship, then there are directed edges (for following) along with undirected edges (for friendship). There is a distinction within the community between the relationships of "friend" and "follower."

Example 12.31 The Job Scheduling Problem
Imagine five jobs (A, B, C, D, E) are to be completed, and suppose they are related by the following mixed graph:



Observe there exist both directed and undirected edges in this graph. The undirected edges indicate temporally independent jobs; that is to say, jobs A and B (as an illustration) can be done at any time (there is no need for simultaneous or chain processes in relation to jobs A and B).

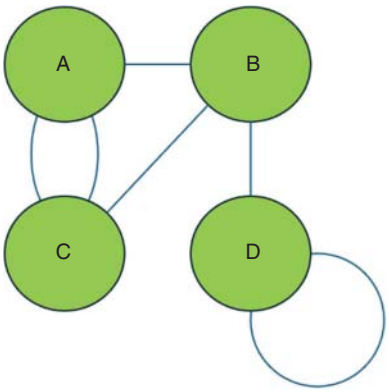
The directed edges indicate the requirement for sequential completion of tasks. For instance, there is a directed edge connecting B to C; only after B is completed can job C be performed. We can interpret this as meaning the output of job B is the input of job C.

Note the processes of jobs A, D, and E are sequential, as are the processes of jobs B and C. Jobs A and B are independent, as are C and D, as well as C and E.

12.3.6 Multigraphs

A multigraph is a type of graph that allows multiple edges between the same pair of vertices. Unlike simple graphs, which have at most one edge between any two vertices, **multigraphs can have multiple edges connecting the same pair of vertices**. Multigraphs can also contain loops, which are edges connecting a vertex to itself.

Example 12.32 Consider a multigraph representing a transportation network between cities, where vertices represent cities, and edges represent different transportation routes, such as roads, trains, or flights.



In this multigraph, the vertices A, B, C, and D represent cities, and the edges represent different transportation routes. There are two edges connecting cities A and C, indicating multiple transportation options between them, such as a road and a train route. The edge connecting vertex D to itself represents a loop, which could signify a local transportation route within the city.

Observe, moreover, there is no direct transportation route from either A or C to D. This brings to mind the notion of travel hubs and spur lines. Taking this view, D is likely some peripheral city within the transportation network; A, B, and C are more likely to be travel hubs, owing to the fact they have routes connecting themselves directly.

Considering the connectivity of B to all three other cities, that location must be a central hub for the transportation network, while A and C are secondary hubs (owing to the fact one cannot travel to ALL cities in the network from either of those cities.

That a single travel line exists from any other city to D, and that D does not connect to any other city, we refer to the connection from B to D as a "spur line," which terminates at D.

Properties of Multigraphs:

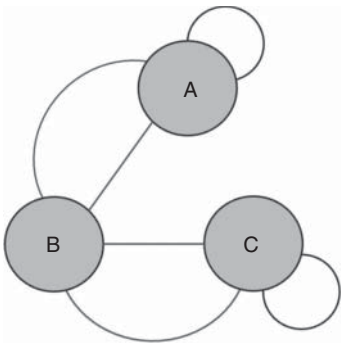
- Multiple Edges: Multigraphs allow multiple edges between the same pair of vertices. These edges are sometimes referred to as parallel edges. Each edge can have its own distinct properties or attributes, such as weight or label.
- Loops: Multigraphs can contain loops, which are edges that connect a vertex to itself. Loops are also known as self-loops or self-edges.*
- Edge Multiplicity: The multiplicity of an edge in a multigraph is the number of occurrences of that edge between the same pair of vertices.
- Degree: The degree of a vertex in a multigraph is the number of edges incident to it. In the presence of loops, each loop contributes twice to the degree of the vertex it is incident to.
- Adjacency: Two vertices in a multigraph are considered adjacent if there is at least one edge connecting them. Multiple edges between the same pair of vertices do not affect the adjacency relationship.

**The issue of loops within a multigraph has no universal acceptance and is a source of contention regarding whether the term "multigraph" should be suppressed altogether. Some authorities require that multigraphs possess no loops, some allow them, and yet others make no reference to acceptance or nonacceptance of the presence of loops.*

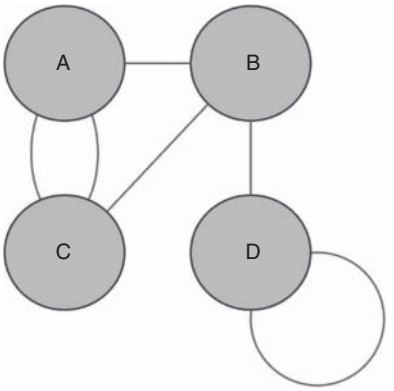
Those who eschew the presence of loops in multigraphs use the term "pseudograph" to refer to a graph permitting both multiple edges AND loops.

*Often, the choice is made whether to permit the existence of loops based on the context of the multigraph. Here, we take no stance on the appropriateness of the terms "multigraph" and "pseudographs" (accepting that both terms are in use) adopt the convention that loops **may exist** within a multigraph.*

Example 12.33 A pseudograph, so named by individuals disallowing the use of loops within a multigraph.



Example 12.34 Identify the multiple edges, loops, edge multiplicities, degrees, and adjacency relationships in the transportation route graph shown:



Solution:
Multiple edges exist connecting city A to city C (hereafter, to avoid repetitive writing, we will omit the word "city" in this explanation).

There is a loop at D.

Edge multiplicity: In the example above, the edge multiplicity between vertices A and C is 2.

Multigraphs have various applications, including:

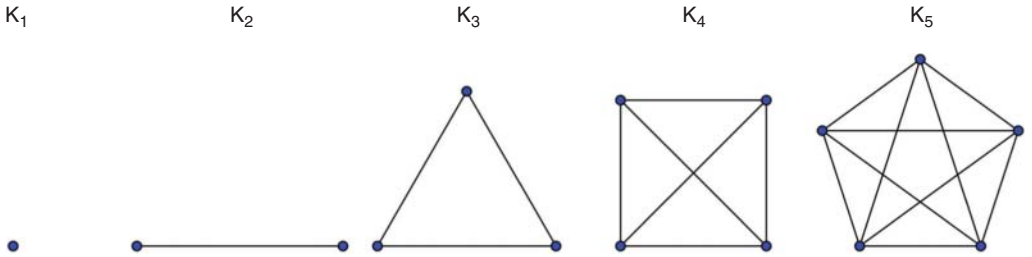
1) Transportation Networks: Multigraphs can model transportation networks, where multiple routes or modes of transportation exist between locations, such as roads, trains, or flights.
2) Electrical Circuits: Multigraphs can represent electrical circuits, where vertices represent components, and edges represent the wires or connections between them. Multiple edges can represent parallel connections or redundant paths.
3) Social Networks: Multigraphs can be used to model social networks with multiple types of relationships between individuals, such as family, friends, or colleagues. Each type of relationship can be represented by a different edge.
4) Dependency Graphs: Multigraphs can represent dependency graphs, where vertices represent tasks or components, and multiple edges represent different types of dependencies or relationships between them.

Multigraphs provide a more flexible and expressive representation compared to simple graphs, allowing for the modeling of systems with multiple connections or parallel relationships between entities. They capture the complexity and richness of real-world networks and enable the analysis of scenarios where multiple edges or loops are significant.

Multigraphs can be further classified into directed multigraphs (where edges have a specific direction) and undirected multigraphs (where edges have no direction). The choice between directed and undirected multigraphs depends on the nature of the relationships being modeled and the requirements of the specific problem or application.

12.3.7 Complete Graphs

A complete graph, denoted by K_n, is a simple graph having n vertices, in which every pair of vertices is connected by a unique edge. In other words, a complete graph has the maximum possible number of edges for a given number of vertices.

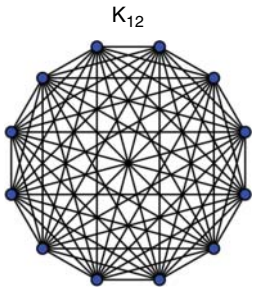


Shown above are representations of the first five complete graphs. Observe the sketches display regularity of shapes, but there is no need for this to be the case.

Properties of Complete Graphs:

- Number of Edges: The number of edges in a complete graph K_n is given by the formula: $n(n-1)/2$. This formula represents the total number of unique pairs of vertices in the graph.
- Degree: In a complete graph K_n, every vertex has a degree of $n-1$. Since every vertex is connected to all other vertices, the degree of each vertex is equal to the total number of vertices minus one.
- Connectivity: Complete graphs are maximally connected. There is a direct edge between every pair of vertices, making the graph highly interconnected. Removing any single vertex or edge does not disconnect the graph.
- Diameter: The diameter of a complete graph, which is the maximum shortest path distance between any two vertices, is always 1. Since every pair of vertices is directly connected by an edge, the shortest path between any two vertices is always of length 1.
- Cliques: A clique is a subgraph where every pair of vertices is connected by an edge. In a complete graph, the entire graph itself is a clique.

For the sample of five complete graphs given earlier, observe the formula for the number of edges holds; in the degenerate case K_1, the number of edges is $\frac{(1)(1-1)}{2} = 0$, for instance. For K_5, the number of edges is $\frac{(5)(5-1)}{2} = \frac{20}{2} = 10$. Using computer assistance, it is a simple matter to sketch K_n rather easily for increasing values of n, but the process becomes unwieldy when done by hand and also visually complex. Here, for example, is K_{12}, which has $\frac{(12)(11)}{2} = 66$ edges:



With the regularity of the shapes, the created pattern is visually elegant, though somewhat challenging to interpret.

Complete graphs have various applications, including:

1) Network Design: Complete graphs can be used to model fully connected networks, such as a small local area network (LAN) where every device is directly connected to every other device, including the server.
2) Social Networks: Complete graphs can represent small social networks or groups where every individual knows every other individual directly.
3) Scheduling Problems: Complete graphs can model scheduling problems where every task or event is related to every other task or event, requiring a complete schedule or timetable.

When solving problems or designing algorithms using complete graphs, it is important to consider the specific properties and characteristics of such graphs. The high connectivity and symmetry of complete graphs can often lead to simplified algorithms or closed-form solutions.

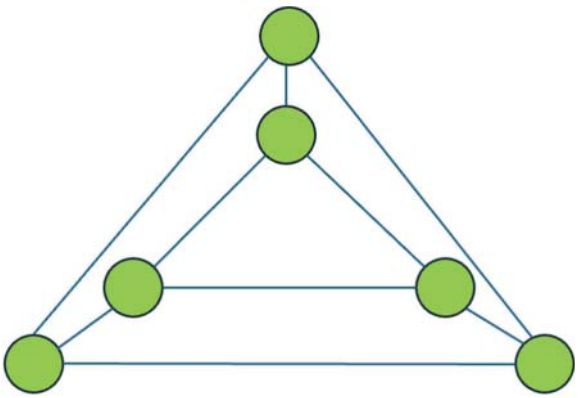
It is worthwhile to contemplate the advantages of the properties of connectivity and symmetry as they pertain to applications.

Connectivity would provide the opportunity for swift communication and an organizational structure that is horizontal, rather than vertical. Though there are advantages to vertical organization of an institution (such as a hierarchy establishing a delineation of authority), use of that hierarchy necessarily creates a time-lag in collaboration and decision-making.

Symmetries in graphs and networks are closely related to the fields of group theory (more specifically, permutation group theory) and graph theory.

Although networks with high levels of symmetry do not often appear among randomly chosen networks, they are often used in practical applications. Networks modeled on vertex-transitive graphs have been shown to be very "good" in their balance of cost (measured by the degree of each vertex in the network) against performance (how easy they are to disconnect, and the efficiency of algorithms run on them).

The notion of "vertex-transitivity" is one we cannot fully address here, but (essentially) a graph is vertex-transitive if there is no structural (i.e., non-labeling) way to distinguish vertices of the graph. In other words, a vertex-transitive graph looks the same from every vertex. An example of this would look like:



Observe each vertex has an identical degree, and so it is impossible, without labeling, to distinguish one vertex from another; we say they are topologically equivalent.

The symmetric properties of a graph are better parameters to measure the stability and synchronizability of an interconnection network.

Complete graphs provide a fundamental building block in graph theory and are used as a reference point for various graph properties and concepts. They represent the extreme case of maximum connectivity and serve as a useful tool for theoretical analysis and problem-solving.

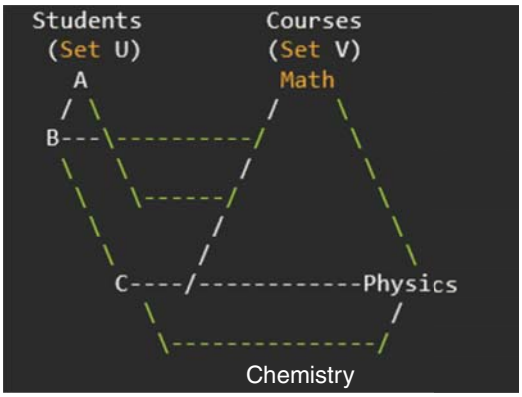
Complete graphs have a high edge density and become impractical for large values of n due to the quadratic growth of the number of edges. In practice, most real-world networks are not complete graphs but exhibit a more complex structure with varying degrees of connectivity between vertices.

12.3.8 Bipartite Graphs

There are many situations where we might want to consider multiple types of objects as vertices, with edges representing asymmetric relationships between two vertices of different types. For example, consider the graph where each vertex is either a house or a person, and two vertices are connected by an edge if one of them is an owner of the other. Amy owns her house, so there is an edge between them, but her house does not own her! This is an example of a **bipartite graph**.

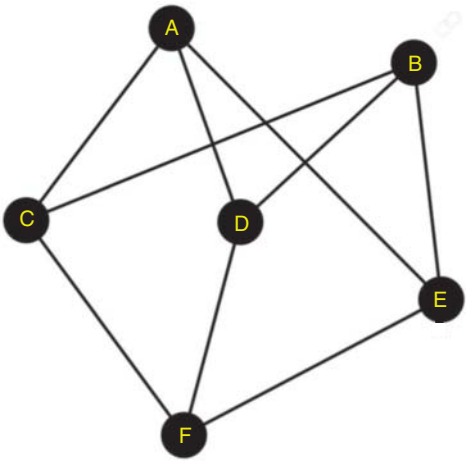
A bipartite graph is a special type of graph in which the vertices can be divided into two disjoint sets, usually denoted as U and V, such that every edge connects a vertex in set U to a vertex in set V. In other words, there are no edges between vertices within the same set. Bipartite graphs are also known as bigraphs.

Example 12.35 Consider a bipartite graph representing the relationship between students and the courses they are enrolled in.

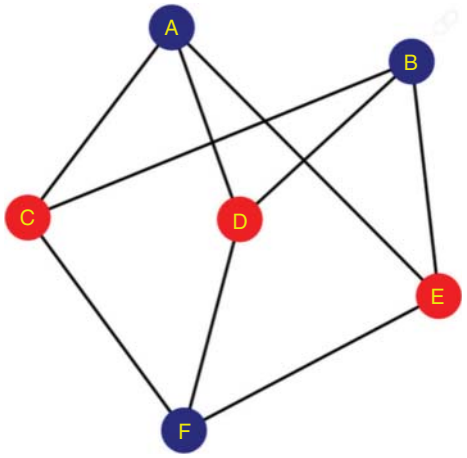


In this bipartite graph, the vertices in set U represent students (A, B, C), and the vertices in set V represent courses (Math, Physics, Chemistry). The edges represent the enrollment of students in specific courses. Student A is enrolled in Math and Physics, student B is enrolled in Math and Chemistry, and student C is enrolled in Physics and Chemistry.

That a graph is bipartite may not be immediately obvious. For example,



In this graph, without more information, we cannot determine which vertices might be grouped together, and so it is impossible to determine whether the graph is bipartite. If, however, we color-code the vertices to indicate which belong to which group,



It becomes evident that the graph is bipartite.

Properties of Bipartite Graphs:

- Vertex Partitioning: The vertices of a bipartite graph can be partitioned into two disjoint sets, U and V, such that every edge connects a vertex from set U to a vertex from set V.
- No Edges within Sets: There are no edges between vertices within the same set. In other words, there are no edges connecting two vertices in set U or two vertices in set V.
- Coloring: Bipartite graphs are two-colorable, meaning that the vertices can be assigned one of two colors such that no two adjacent vertices have the same color. One color can be assigned to vertices in set U, and the other color can be assigned to vertices in set V. See the previous example as an illustration of this property.

- Cycles: If a bipartite graph contains a cycle, the cycle must have an even length. This is because the vertices alternate between sets U and V along the cycle, and the cycle must end at the same set where it started.
- Matching: A matching in a bipartite graph is a subset of edges such that no two edges share a common vertex. Finding maximum matchings in bipartite graphs is a well-studied problem with applications in various domains.

Bipartite graphs have numerous applications, including:

1) Job Assignments: Bipartite graphs can model job assignment problems, where one set of vertices represents workers, and the other set represents tasks. The edges represent the compatibility or skills of workers for specific tasks.
2) Recommender Systems: Bipartite graphs can be used in recommender systems, where one set of vertices represents users, and the other set represents items (e.g., movies and products). The edges represent user preferences or ratings for specific items.
3) Chemical Reactions: Bipartite graphs can model chemical reactions, where one set of vertices represents reactants, and the other set represents products. The edges represent the participation of reactants in specific reactions.
4) Scheduling: Bipartite graphs can be used in scheduling problems, where one set of vertices represents time slots, and the other set represents tasks or events. The edges represent the assignment of tasks to specific time slots.

When solving problems or designing algorithms for bipartite graphs, it is important to consider their specific properties and characteristics. The bipartite structure allows for specialized algorithms and optimizations, such as the Hungarian algorithm for finding maximum matchings.

Bipartite graphs provide a useful framework for modeling and analyzing relationships between two distinct sets of entities. They capture the connections or associations between elements of different types and enable the study of various optimization and matching problems.

Not all graphs are bipartite. Graphs that cannot be partitioned into two disjoint sets satisfying the bipartite property are called non-bipartite graphs. The ability to determine whether a graph is bipartite or not is an important problem in graph theory and has implications for various applications and algorithms.

12.3.9 Regular Graphs

A regular graph is a graph in which all vertices have the same degree. In other words, every vertex in a regular graph is connected to the same number of other vertices. The degree of a vertex in a regular graph is often denoted as r, and the graph is called an r-regular graph.

Properties of Regular Graphs:

- Degree: In an r-regular graph, every vertex has a degree of r. The degree of a vertex represents the number of edges incident to that vertex.
- Number of Edges: The total number of edges in an r-regular graph with n vertices is given by the formula: (n * r)/2. This formula holds because each edge is counted twice (once for each endpoint), and the total degree of all vertices is equal to n * r.
- Connectivity: Regular graphs can be connected or disconnected. A connected regular graph has a single component, while a disconnected regular graph consists of multiple components, each of which is a regular graph.

- Completeness: A complete graph is a special case of a regular graph where every vertex is connected to every other vertex. In a complete graph with n vertices, each vertex has a degree of $n - 1$, making it an $(n - 1)$-regular graph.
- Cycles: Regular graphs can contain cycles. In fact, cycle graphs are a type of regular graph where each vertex has a degree of 2.

Regular graphs have various applications, including:

1) Network Topology: Regular graphs can be used to model network topologies where each node has the same number of connections. For example, a four-regular graph can represent a network where each device is connected to four other devices.
2) Load Balancing: Regular graphs can be used in load balancing scenarios where tasks or resources need to be evenly distributed among nodes. Each node in a regular graph has the same capacity or responsibility.
3) Cryptography: Regular graphs, particularly expander graphs, have applications in cryptography. Expander graphs have the property that every small subset of vertices has a relatively large number of neighbors, making them useful for constructing secure communication networks.
4) Error-Correcting Codes: Regular graphs, such as bipartite graphs, are used in the construction of error-correcting codes. The regularity property helps in achieving good encoding and decoding properties.

Regular graphs provide a structured and balanced representation of networks or systems where each entity has the same level of connectivity or involvement. They offer a foundation for studying various graph-theoretic concepts and have practical applications in fields such as network design, load balancing, and cryptography.

Graphs with varying degrees among vertices are called **irregular graphs**.

12.4 Adjacency Matrix

An adjacency matrix is a square matrix used to represent a finite graph. It is a two-dimensional array that indicates the presence or absence of edges between vertices in the graph. The size of the adjacency matrix is equal to the number of vertices in the graph.

Example 12.36 Consider the following undirected graph:



The adjacency matrix representation of this graph is:



In this adjacency matrix, the rows and columns represent the vertices of the graph (A, B, C, D, E). The entries in the matrix indicate the presence or absence of edges between the corresponding vertices. A value of 1 in the matrix indicates the presence of an edge, while a value of 0 indicates the absence of an edge.

For example, the entry in row A and column B is 1, indicating that there is an edge between vertices A and B. Similarly, the entry in row C and column D is 1, indicating an edge between vertices C and D. The diagonal entries of the matrix are always 0 since there are no self-loops in this graph.

Properties of Adjacency Matrix:

- Symmetry: For an undirected graph, the adjacency matrix is symmetric. The entry in row *i* and column *j* is the same as the entry in row *j* and column *i*. This is because an edge between vertices *i* and *j* is the same as an edge between vertices *j* and *i*.
- Diagonal Entries: In a simple graph without self-loops, the diagonal entries of the adjacency matrix are always 0. This is because, in a simple graph, a vertex cannot have an edge to itself.
- Space Complexity: The space complexity of an adjacency matrix is $O(V^2)$, where V is the number of vertices in the graph. This is because the matrix requires $V \times V$ space to store the edge information.
- Edge Queries: Checking the presence or absence of an edge between two vertices can be done in constant time O(1), by accessing the corresponding entry in the adjacency matrix.
- Weighted Graphs: For weighted graphs, the adjacency matrix can store the weights of the edges instead of just 1s and 0s. The entry in row *i* and column *j* represents the weight of the edge between vertices *i* and *j*.

Adjacency matrices have various applications, including:

1) Graph Algorithms: Adjacency matrices are used in various graph algorithms, such as depth-first search (DFS), breadth-first search (BFS), and shortest path algorithms like Floyd–Warshall.
2) Network Analysis: Adjacency matrices can represent social networks, communication networks, or any other network where the relationships between entities can be represented as a graph.
3) Optimization Problems: Adjacency matrices are used in solving optimization problems on graphs, such as finding the maximum flow or minimum cut in a network.

When using an adjacency matrix representation, it is important to consider the trade-offs. While adjacency matrices provide fast edge queries and simplify certain algorithms, they can be space-inefficient for sparse graphs (graphs with few edges compared to the number of vertices). In such cases, **adjacency lists** (which we shall introduce momentarily) might be more suitable.

Adjacency matrices can be modified to represent directed graphs by allowing asymmetry in the matrix entries. In a directed graph, the entry in row *i* and column *j* represents an edge from vertex *i* to vertex *j*, but the entry in row *j* and column *i* may not necessarily represent an edge from vertex *j* to vertex *i*.

12.4.1 Adjacency List

An adjacency list is a way to represent a graph using a collection of lists. Each vertex in the graph has its own list that contains the vertices adjacent to it. The adjacency list representation is particularly

useful for sparse graphs (graphs with few edges compared to the number of vertices) as it provides a space-efficient way to store the graph.

Example 12.37 Consider the following undirected graph:



The adjacency list representation of this graph is:

A: [B, C, D]
B: [A, E]
C: [A, D]
D: [A, C, E]
E: [B, D]

In this adjacency list, each vertex (A, B, C, D, E) has its own list that contains the vertices it is connected to. For example, vertex A is connected to vertices B, C, and D, so its adjacency list is [B, C, D]. Similarly, vertex B is connected to vertices A and E, so its adjacency list is [A, E].

Properties of Adjacency List:

- Space Complexity: The space complexity of an adjacency list is O(V + E), where V is the number of vertices and E is the number of edges in the graph. This is because each vertex has a list of its adjacent vertices, and the total number of entries in all the lists is equal to the number of edges.
- Edge Queries: Checking the presence or absence of an edge between two vertices requires traversing the adjacency list of one of the vertices. The time complexity of this operation is O(d), where d is the degree of the vertex (the number of vertices adjacent to it).
- Vertex Queries: Accessing the neighbors of a vertex can be done efficiently by traversing its adjacency list. The time complexity of this operation is O(d), where d is the degree of the vertex.
- Weighted Graphs: For weighted graphs, the adjacency list can be modified to store the weights along with the adjacent vertices. Each entry in the adjacency list becomes a pair (vertex, weight) representing the adjacent vertex and the weight of the edge.
- Directed Graphs: Adjacency lists can easily represent directed graphs. In a directed graph, each vertex's adjacency list contains only the vertices that are directly reachable from it, following the direction of the edges.

Adjacency lists have various applications, including:

1) Graph Traversal: Adjacency lists are commonly used in graph traversal algorithms such as DFS and BFS. The adjacency lists provide efficient access to the neighbors of each vertex during the traversal.
2) Shortest Path Algorithms: Adjacency lists are used in shortest path algorithms like Dijkstra's algorithm and Bellman–Ford algorithm. The adjacency lists allow efficient exploration of the neighboring vertices and their distances.
3) Network Routing: Adjacency lists can represent network topologies and assist in routing algorithms. Each vertex represents a network node, and the adjacency lists indicate the connections between nodes.

The implementation of adjacency lists can vary depending on the programming language and data structures used. Common implementations include using an array or list of lists or using a hash table to store the adjacency lists for each vertex.

12.4.2 Incidence Matrix

An incidence matrix is a way to represent a graph using a matrix that shows the relationship between vertices and edges. It is a two-dimensional matrix where the rows represent the vertices of the graph, and the columns represent the edges. The entries in the matrix indicate whether a vertex is incident to an edge or not.

Properties of Incidence Matrix:

- Matrix Size: The size of the incidence matrix is $V \times E$, where V is the number of vertices and E is the number of edges in the graph.
- Row Sum: In an undirected graph, the sum of the entries in each row of the incidence matrix is equal to the degree of the corresponding vertex. This is because each edge is incident to two vertices, and each vertex is incident to the edges connected to it.
- Column Sum: In an undirected graph, the sum of the entries in each column of the incidence matrix is always 2. This is because each edge is incident to exactly two vertices.
- Directed Graphs: For directed graphs, the incidence matrix can be modified to indicate the direction of the edges. A value of 1 can represent an outgoing edge, while a value of -1 can represent an incoming edge.
- Weighted Graphs: Incidence matrices can be extended to represent weighted graphs by storing the weights of the edges instead of just 1s and 0s. The entries in the matrix can represent the weights of the corresponding edges.

Incidence matrices have various applications, including:

1) Graph Visualization: Incidence matrices provide a compact representation of the graph structure and can be used for graph visualization and analysis.
2) Network Flow Problems: Incidence matrices are used in solving network flow problems, such as maximum flow and minimum cut problems. The matrix representation allows for efficient computation of flow constraints.
3) Electrical Networks: Incidence matrices can represent electrical networks, where vertices represent nodes and edges represent branches. The matrix helps in analyzing the connectivity and current flow in the network.

Incidence matrices are not as commonly used as adjacency matrices or adjacency lists due to their higher space complexity. However, they find applications in specific graph problems where the incidence relationships are crucial, such as network flow problems and electrical network analysis.

12.4.3 Edge List

An edge list represents a graph by listing all the edges in the graph. It is a collection or array of pairs, where each pair represents an edge connecting two vertices. The edge list representation is particularly useful for sparse graphs (graphs with few edges compared to the number of vertices) and when the focus is on the edges rather than the vertices.

Example 12.38 Consider the following undirected graph:



The edge list representation of this graph is:

[(A, B), (A, C), (A, D), (B, E), (C, D), (D, E)]

In this edge list, each pair represents an edge in the graph. For example, the pair (A, B) represents the edge connecting vertices A and B. Similarly, the pair (C, D) represents the edge connecting *vertices C and D*.

Properties of Edge Lists:

- Space Complexity: The space complexity of an edge list is O(E), where E is the number of edges in the graph. This is because each edge is stored as a pair in the list.
- Vertex Queries: Finding the neighbors of a vertex requires traversing the entire edge list and selecting the edges that contain the vertex. The time complexity of this operation is O(E), as it may require examining all the edges in the worst case.
- Edge Queries: Checking the presence or absence of an edge between two vertices can be done by searching the edge list for the corresponding pair. The time complexity of this operation is O(E), as it may require searching through all the edges in the worst case.
- Weighted Graphs: For weighted graphs, the edge list can be extended to include the weights along with the vertex pairs. Each entry in the edge list becomes a tuple (u, v, w), where u and v are the vertices, and w is the weight of the edge.
- Directed Graphs: Edge lists can represent directed graphs by using ordered pairs to indicate the direction of the edges. For example, the pair (u, v) represents an edge from vertex u to vertex v.

Edge lists have various applications, including:

1) Graph Construction: Edge lists are often used as an input format for constructing graphs. They provide a compact and easy-to-read representation of the edges in the graph.
2) Graph Algorithms: Edge lists can be used in graph algorithms that primarily focus on the edges, such as Kruskal's algorithm for finding MSTs or algorithms for finding bridges or articulation points in a graph.
3) Data Analysis: Edge lists are commonly used in data analysis tasks, such as analyzing social networks or recommendation systems. They provide a simple representation of the relationships between entities.

Edge lists can be combined with other graph representations, such as adjacency lists or adjacency matrices, to provide a more comprehensive representation of the graph. The choice of representation depends on the specific requirements of the problem and the operations that need to be performed on the graph.

12.5 Graph Traversals

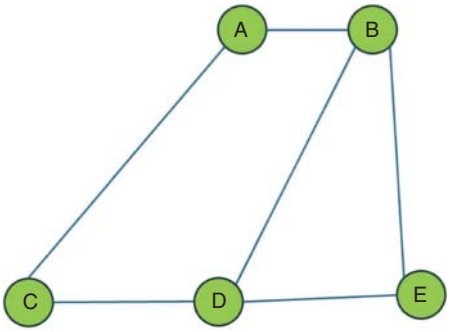
Graph traversal is the process of visiting all the vertices in a graph in a particular order. It is a fundamental operation in graph theory and forms the basis for many algorithms and their applications. In this section, we will discuss one of the most common graph traversal techniques: DFS.

12.5.1 Depth-First Search (DFS)

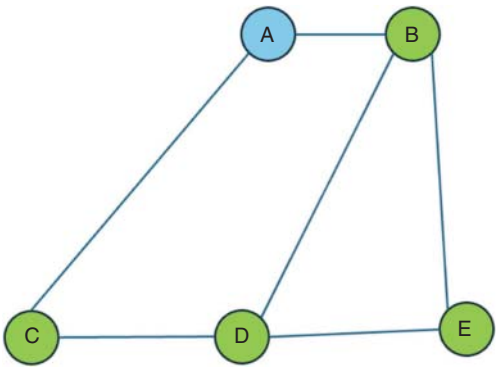
DFS is a recursive graph traversal algorithm (used for both directed and undirected graphs) that explores the graph beginning at a "source vertex." The method proceeds by visiting vertices as far along a series of edges as possible along each investigative branch before backtracking. By "branch," we mean a curve drawn between two vertices to indicate a connection between those vertices. More technically, a **tree** is a connected graph without any cycles, and the edges of a tree are called **branches**.

12.5.1.1 Algorithm

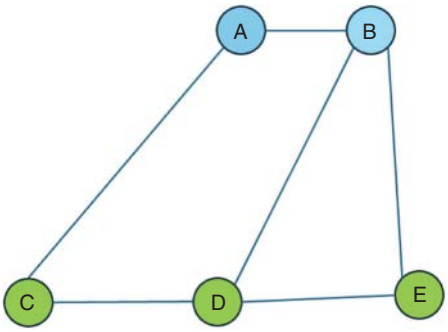
The DFS algorithm will be explained and demonstrated, working with the following undirected graph:



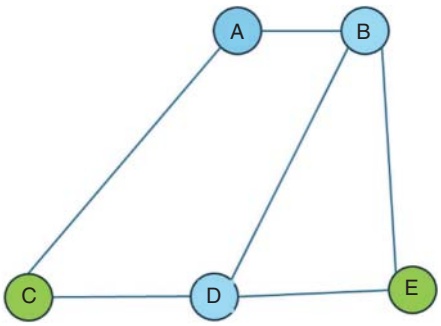
- Choose a starting vertex and mark it (in some manner) as "visited." Randomly, we will choose to start at vertex A; we will mark it as "visited" by changing the color to blue.



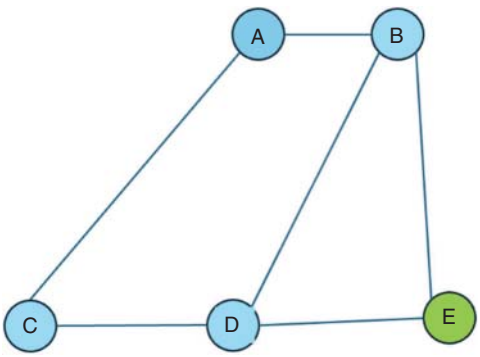
- Explore an unvisited adjacent vertex by marking it as visited. We will progress alphabetically, but there is no reason to do so other than a personal choice.



3. From B, the adjacent vertices are D and E; we choose to visit D.



The adjacent vertices to D are C and E; we choose to visit C.

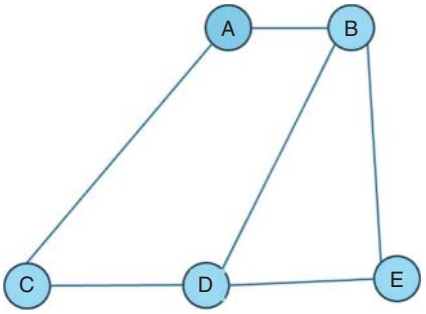


- If all adjacent vertices have been visited, backtrack to the previous vertex.

  We return to vertex D.

- Repeat steps 2 and 3 until all reachable vertices have been visited.

The vertex adjacent to D that has not been visited is E, which we now visit. This achieves a visitation of all vertices.



The DFS traversal order for starting at A is: A–B–D–C–D–E.

The time complexity of the DFS algorithm is O(V + E), where V is the number of vertices, and E is the number of edges in the graph. This is because DFS visits each vertex once and explores each edge once.

12.5.1.2 Applications

DFS has applications in graph theory and computer science, including:

1) Connectivity: DFS can be used to determine the connectivity of a graph. If DFS can reach all vertices from a given starting vertex, the graph is connected.
2) Cycle Detection: DFS can be used to detect cycles in a graph. If during the DFS traversal, we encounter a vertex that has already been visited and is not the parent of the current vertex, then there exists a cycle in the graph.
3) Topological Sorting: DFS can be used to perform topological sorting of a directed acyclic graph (DAG). Topological sorting is an ordering of the vertices such that for every directed edge from vertex A to vertex B, vertex A comes before vertex B in the ordering.
4) Strongly Connected Components: DFS can be used to find SCCs in a directed graph. An SCC is a maximal subgraph in which there is a path from every vertex to every other vertex.
5) Solving Mazes and Puzzles: DFS can be used to solve mazes and puzzles by exploring all possible paths from a starting point until the goal is reached or all paths have been exhausted.

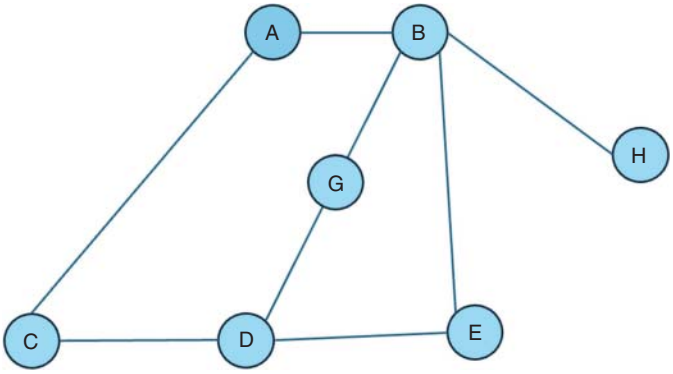
DFS can be implemented using different data structures, such as recursion or an explicit stack. The choice of implementation depends on the specific requirements of the problem and the programming language being used.

12.5.2 Breadth-First Search (BFS)

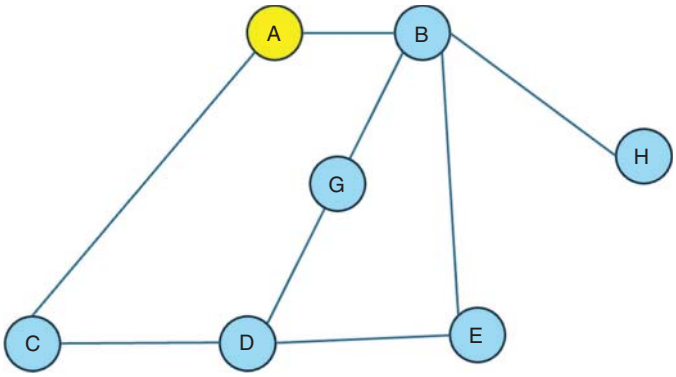
BFS is another graph traversal algorithm that explores a graph by visiting all neighbors equidistant from the starting vertex before moving on to the next level of vertices. Unlike DFS, which explores as far as possible along each branch, BFS explores the graph level by level, visiting all the vertices at the current level before moving on to the next level.

In a sense, BFS is examining a hierarchical structure, considering all vertices at a particular distance from the starting vertex before moving on to those at the next distance.

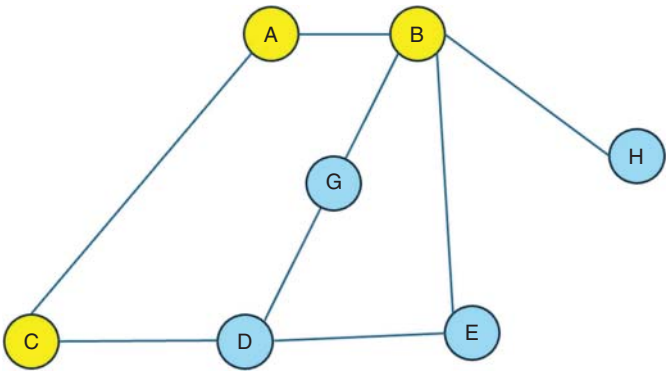
Let us perform the BFS on a graph to illustrate the technique.



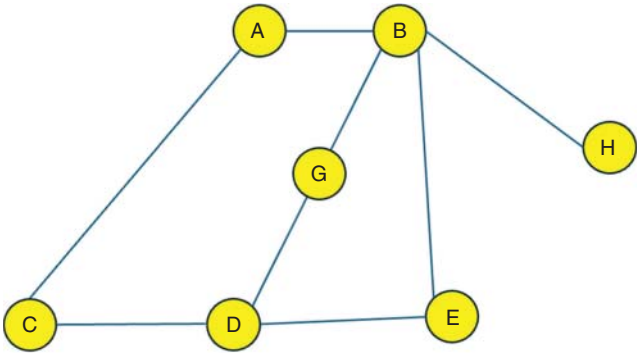
We will start at vertex A, changing its color to yellow to signify it has been visited.



Having identified a particular starting vertex, we now visit (in whichever order we want – BFS will yield different, equally valid search patterns depending upon our choices) those vertices which are at distance 1 from vertex A. The two vertices we must examine next are B and C; to be "different," we will choose C first, then B. Each of these is now marked as having been visited.

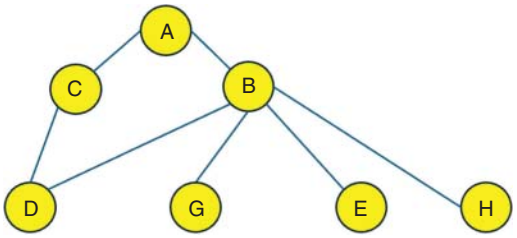


We now consider which vertices are at distance 2 from our starting vertex A. These would be G, H, E, and D.



Having visited all the vertices of the graph, we have completed the BFS search algorithm. Shown in order of visitation, our result is A–C–B–G–H–E–D.

It is natural to wonder how the notion of a hierarchy appears in this example. If we reorient the graph so the vertices equidistant from the starting vertex are aligned horizontally (while preserving their edge connectivity), we see:



It is here, by the way, where we see the inspiration for the name "Breadth-First." The word "breadth" refers to "width," and what we are doing is progressing from level to level in the graph, covering the entire width of that level in our search, before moving on.

The time complexity of the BFS algorithm is O(V + E), where V is the number of vertices, and E is the number of edges in the graph. This is because BFS visits each vertex once and explores each edge once.

12.5.2.1 Applications

BFS has several applications in graph theory and computer science, including:

1) Shortest Path: BFS can be used to find the shortest path between two vertices in an unweighted graph. The shortest path is the path with the minimum number of edges between the source and destination vertices.

NOTE: BFS alone does not identify the shortest path, because BFS is not looking for a shortest path: it merely describes a strategy for searching a graph, but it does not say that you must search for anything in particular.

2) Connected Components: BFS can be used to find connected components in an undirected graph. A connected component is a subgraph in which there is a path between every pair of vertices.
3) Bipartiteness Testing: BFS can be used to test whether a graph is bipartite or not. A bipartite graph is a graph whose vertices can be divided into two disjoint sets such that every edge connects a vertex in one set to a vertex in the other set.

4) Level-Order Traversal: BFS can be used to perform a level-order traversal of a tree or a graph. In a level-order traversal, the vertices are visited level by level, from the root or starting vertex down to the leaves or farthest vertices.
5) Network Analysis: BFS can be used in network analysis to find the shortest paths, determine the connectivity of a network, or identify the most influential nodes based on their proximity to other nodes.

"Web crawlers" (bots typically incorporated within search engines, meant to index the content of websites all across the Internet so that those websites can appear in search engine results) use BFS to systematically explore the web.

BFS starts from a set of seed URLs and visits all reachable pages, layer by layer. This ensures that the crawler covers all pages within a certain depth before moving deeper into the web.

Search engines like Google employ BFS-based web crawlers to index web pages as efficiently as is practical, which assists in maintaining both freshness and comprehensiveness of search engine databases.

When constructing a program to perform BFS, we require an additional data structure, typically a queue, to keep track of the vertices to be visited. The choice of data structure and implementation may vary depending on the programming language and specific requirements of the problem.

12.6 Connectivity

12.6.1 Connected Graphs

A graph is said to be **connected** if there exists a path between every pair of vertices in the graph. In other words, for any two vertices in the graph, there is a sequence of edges that allows you to travel from one vertex to the other. If a graph is not connected, it is referred to as a **disconnected graph**.

A Connected Graph:



A Disconnected Graph:



Graph 1 is a connected graph because there exists a path between every pair of vertices. For example:

a) To go from vertex A to vertex E, you can follow the path: A –> B –> E.
b) To go from vertex C to vertex D, you can follow the path: C –> D.
c) To go from vertex A to vertex D, you can follow the path: A –> C –> D or A –> D.

On the other hand, Graph 2 is a disconnected graph because there is no path between certain pairs of vertices. In this case, there is no path between any vertex in the subgraph {A, B} and any vertex in the subgraph {C, D}. The graph consists of two separate connected components.

Properties of connected graphs:

1) Reachability: In a connected graph, every vertex is reachable from every other vertex. This means that there exists a path between any two vertices in the graph.
2) Single connected component: A connected graph consists of a single connected component. A connected component is a maximal subgraph in which there is a path between every pair of vertices.
3) Removal of edges: If the removal of an edge from a connected graph results in a disconnected graph, then that edge is called a **bridge**. A bridge is an edge whose removal increases the number of connected components in the graph.
4) Removal of vertices: If the removal of a vertex (and all its incident edges) from a connected graph results in a disconnected graph, then that vertex is called an **articulation point** or a **cut vertex**. An articulation point is a vertex whose removal increases the number of connected components in the graph.

Connectivity has several applications in graph theory and real-world scenarios, including:

1) Network Reliability: Connectivity is essential in network design and analysis. A connected network ensures that there is a path between any two nodes, enabling communication and data transmission. Network administrators aim to maintain connectivity and identify potential points of failure (bridges or articulation points) to enhance network reliability.
2) Graph Traversal: Connectivity is a prerequisite for graph traversal algorithms such as DFS and BFS. These algorithms rely on the existence of paths between vertices to explore and process the graph.
3) Clustering and Community Detection: Connectivity is used in cluster analysis and community detection algorithms. Connected components in a graph can represent clusters or communities of related entities, helping in data analysis and pattern recognition.
4) Shortest Path and Distance: Connectivity allows for the calculation of shortest paths and distances between vertices in a graph. Algorithms like Dijkstra's shortest path algorithm and Floyd–Warshall algorithm require the graph to be connected to determine the minimum distance or shortest path between vertices.

Properties of disconnected graphs:

1) Multiple connected components: A disconnected graph consists of two or more connected components. Each connected component is a maximal subgraph in which there is a path between every pair of vertices within that component.
2) Unreachable vertices: In a disconnected graph, there exist pairs of vertices for which there is no path connecting them. Vertices in different connected components are unreachable from each other.
3) Isolated vertices: An isolated vertex is a vertex that has no incident edges. In a disconnected graph, an isolated vertex forms its own connected component.
4) Bridges and articulation points: In a disconnected graph, bridges and articulation points do not exist. Bridges and articulation points are relevant only in connected graphs, where their removal increases the number of connected components.

Disconnected graphs have several applications and considerations, including:

1) Network Segmentation: Disconnected graphs can represent segmented or partitioned networks. Each connected component can represent a separate subnetwork or an isolated group of nodes that are not reachable from other parts of the network.

2) Graph Decomposition: Disconnected graphs can be decomposed into their connected components. Each connected component can be analyzed and processed independently, simplifying certain graph algorithms and operations.
3) Graph Visualization: When visualizing disconnected graphs, it is common to represent each connected component separately. This helps in identifying the isolated subgraphs and understanding the overall structure of the graph.
4) Connectivity Testing: To determine whether a graph is connected or disconnected, algorithms like DFS or BFS can be used. By starting a traversal from any vertex and checking if all vertices are visited, the connectivity of the graph can be determined.
5) Graph Algorithms: Some graph algorithms may need to handle disconnected graphs differently. For example, algorithms for finding shortest paths or spanning trees may need to be adapted to work on each connected component separately.

Disconnected graphs provide a way to represent and analyze graphs that are not fully interconnected. They highlight the presence of separate subgraphs or isolated components within a graph.

Disconnected graphs can be transformed into connected graphs by adding bridges between the connected components. This process is known as **connecting** or **bridging the components**. However, the choice to connect components depends on the specific problem or application at hand.

12.6.2 Strongly Connected Graphs

In directed graphs, the concept of connectivity is further refined into **strong connectivity**. A directed graph is said to be strongly connected if there is a directed path from every vertex to every other vertex in the graph. In other words, for any pair of vertices u and v, there exists a directed path from u to v and a directed path from v to u.

Here's a detailed example to illustrate strongly connected graphs:

Consider the following directed graph:



This graph is strongly connected because there is a directed path between every pair of vertices:

1) From A to B : A –> B.
2) From B to A : B –> D –> C –> A.
3) From A to C : A –> B –> D –> C.
4) From C to A : C –> A.
5) From B to C : B –> D –> C.
6) From C to B : C –> A –> B.
7) From D to A : D –> C –> A.
8) From A to D : A –> B –> D.

- An SCC is a maximal subgraph of a directed graph in which there is a directed path from every vertex to every other vertex within the subgraph. In a strongly connected graph, the entire graph itself is a single SCC.

- Condensation graph: The condensation graph of a directed graph is obtained by collapsing each SCC into a single vertex. The resulting graph is a DAG, where each vertex represents an SCC of the original graph.
- Algorithms for finding SCCs: There are efficient algorithms for finding SCCs in a directed graph, such as Kosaraju's algorithm and Tarjan's algorithm. These algorithms have a time complexity of $O(V + E)$, where V is the number of vertices, and E is the number of edges in the graph.

Strongly connected graphs have several applications, including:

1) Communication Networks: In communication networks, strong connectivity ensures that there is a directed path for information to flow from any node to any other node in the network. This is crucial for reliable communication and data transmission.
2) Web Graph Analysis: The World Wide Web can be modeled as a directed graph, where web pages are vertices and hyperlinks are directed edges. SCCs in the web graph can represent closely related groups of web pages or communities with high interconnectivity.
3) Social Network Analysis: In social networks, strong connectivity can indicate tight-knit groups or communities where individuals have mutual connections and can reach each other through a sequence of relationships.
4) Recommendation Systems: SCCs in a recommendation graph can represent groups of items or users with similar preferences or interests. This information can be used to generate personalized recommendations.
5) Scheduling and Dependency Analysis: In scheduling problems or dependency analysis, strong connectivity can represent tasks or events that are interdependent and require a specific order of execution.

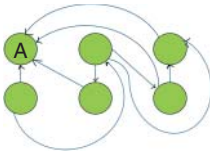
A directed graph can have multiple SCCs. The condensation graph, obtained by collapsing each SCC into a single vertex, provides a high-level view of the connectivity structure of the graph. The condensation graph is always a DAG, as there cannot be any cycles between SCCs.

12.6.3 Weakly Connected Graphs

In directed graphs, another notion of connectivity is **weak connectivity**. A directed graph is said to be weakly connected if its underlying *undirected* graph is connected. In other words, if we disregard the direction of the edges and treat the graph as an undirected graph, it becomes connected.

Here's a detailed example to illustrate weakly connected graphs:

Consider the following directed graph:



Notice there is no way to get from the vertex labeled as "A" to any other vertex. This digraph is NOT strongly connected. However, if we disregard the orientation of the edges, the new graph (the one underlying the directed graph) IS connected.

Therefore, this directed graph is weakly connected.

Properties of weakly connected graphs:

1) A directed graph is weakly connected if its underlying undirected graph, obtained by disregarding the direction of the edges, is connected.
2) Every strongly connected graph is also weakly connected, but the converse is not necessarily true. A weakly connected graph may not be strongly connected.
3) In a weakly connected graph, the connected components of the underlying undirected graph correspond to the weakly connected components of the directed graph.
4) In a weakly connected graph, there may be pairs of vertices for which there is no directed path from one vertex to the other. However, there exists an undirected path between every pair of vertices.

Weakly connected graphs have several applications and considerations, including:

1) Transportation Networks: In transportation networks, weak connectivity can represent the overall reachability of locations, even if some routes are one-way or have restrictions. As long as there is an undirected path between any two locations, the network is considered weakly connected.
2) Information Flow: Weakly connected graphs can model information flow in scenarios where the direction of the flow may not be strictly enforced. For example, in a social network, information can spread between individuals even if the relationships are not reciprocal.
3) Graph Exploration: Weak connectivity is relevant in graph exploration algorithms where the goal is to visit all vertices, regardless of the edge directions. Algorithms like DFS or BFS can be used to explore weakly connected graphs.
4) Connectivity Testing: To determine whether a directed graph is weakly connected, the underlying undirected graph can be constructed, and connectivity testing algorithms can be applied to the undirected graph.

Weakly connected graphs provide a way to analyze connectivity in directed graphs by considering the underlying undirected structure. They have applications in transportation networks, information flow, and graph exploration.

12.6.4 Components

The term "component" refers to a maximal connected subgraph of an undirected graph or a maximal strongly connected subgraph of a directed graph.

Connected Components in Undirected Graphs

In an undirected graph, a connected component is a maximal subgraph in which there is a path between every pair of vertices. In other words, it is a subgraph that is connected and cannot be expanded further without losing connectivity.

Here is an example of connected components in an undirected graph:



This graph has three connected components:

1) Component 1: {A, B}.
2) Component 2: {C, D}.
3) Component 3: {E, F}.

Each connected component is a maximal connected subgraph. There is no path between vertices in different components.

Properties of connected components in undirected graphs:

1) Within each connected component, there is a path between every pair of vertices.
2) A connected component is maximal, meaning it cannot be expanded further without losing connectivity. Adding any vertex or edge from outside the component would result in a larger connected subgraph.
3) Connected components are disjoint, meaning they do not share any vertices or edges. Each vertex belongs to exactly one connected component.
4) The number of connected components in an undirected graph can be determined by performing a graph traversal algorithm, such as DFS or BFS, and counting the number of times the traversal is initiated.

Strongly Connected Components in Directed Graphs:

In a directed graph, an SCC is a maximal subgraph in which there is a directed path from every vertex to every other vertex within the subgraph.

12.7 Special Graphs

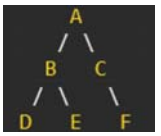
There are several types of graphs that possess unique properties and have significant applications. These special graphs exhibit a specific structure or satisfy certain constraints. In this section, we will explore one such special graph: **trees**.

12.7.1 Trees

12.7.1.1 Definition

A tree is a connected undirected graph that has no cycles. In other words, a tree is a graph in which any two vertices are connected by exactly one path. Trees have a hierarchical structure and are widely used in various fields, including computer science, data structures, corporate organization, and network topology.

Here's an example of a tree:



Observe vertex A is connected to vertices B and C, vertex B is connected to vertices D and E, and vertex C is connected to vertex F. There are no cycles in the graph, and there is exactly one path between any two vertices; therefore, this is a tree.

12.7.1.2 Properties

Trees have several important properties that distinguish them from other graphs:

1) Connectivity: A tree is a connected graph, meaning there is a path between every pair of vertices.
2) No Cycles: A tree does not contain any cycles. If a graph has a cycle, it cannot be a tree.

3) Unique Path: In a tree, there is exactly one path between any two vertices. This property ensures that there is no ambiguity in traversing the tree from one vertex to another.
4) Number of Edges: In a tree with n vertices, there are exactly $n - 1$ edges. This property holds because adding one more edge to a tree would create a cycle, violating the definition of a tree.
5) Acyclic: Trees are acyclic graphs, meaning they do not contain any cycles. Removing any edge from a tree will disconnect the graph, creating two separate trees.
6) Rooted Trees: A tree can be designated as a rooted tree by selecting one vertex as the root. The root serves as the topmost vertex in the hierarchical structure, and every other vertex has a unique path to the root.
7) Parent–Child Relationship: In a rooted tree, each vertex (except the root) has exactly one parent vertex and zero or more child vertices. The parent–child relationship defines the hierarchy of the tree.
8) Leaf Nodes: Vertices in a tree that have no children are called leaf nodes or leaves. In the example tree above, vertices D, E, and F are leaf nodes.

Consider the following tree:



1) The tree is connected, as there is a path between every pair of vertices.
2) There are no cycles in the tree.
3) There is exactly one path between any two vertices. For example, the path from vertex D to vertex F is D – B – A – C – F.
4) The tree has seven vertices ($n = 7$) and six edges ($n-1 = 6$).
5) Vertex A can be considered as the root of the tree.
6) Vertex B is the parent of vertices D and E, while vertex E is the parent of vertex G.
7) Vertices D, G, and F are leaf nodes.

12.7.2 Planar Graphs

12.7.2.1 Definition

A planar graph is a graph that can be drawn on a plane (or a two-dimensional surface) without any edges crossing each other. In other words, it is possible to embed the graph in the plane so that no two edges intersect, except possibly at their endpoints (vertices).

Here's an example of a planar graph:



In this graph, all the edges can be drawn without any crossings. The edges {A, B}, {A, C}, {A, D}, {B, E}, {C, D}, and {D, E} do not intersect with each other, except at their endpoints.

On the other hand, here's an example of a nonplanar graph (the complete graph K_5):



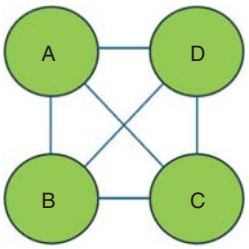
In this graph, it is impossible to draw the edges without any crossings. The edge {A, D} must intersect with either {B, C} or {B, E}, and the edge {B, C} must intersect with either {A, E} or {D, E}.

Planar graphs have several interesting properties and have applications in various fields, including circuit design, geographic information systems, and computer graphics.

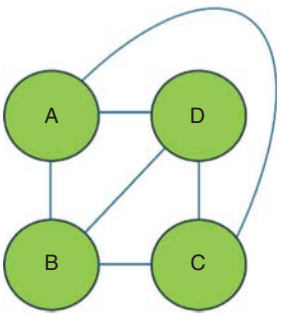
12.7.2.2 Euler's Formula

Euler's formula, named after the famous Swiss mathematician Leonhard Euler (1707–1783), is a fundamental relation in planar graph theory that connects the number of vertices (V), edges (E), and faces (F) in a connected planar graph. Before stating the result, we must first specify what is meant by a face in a planar graph.

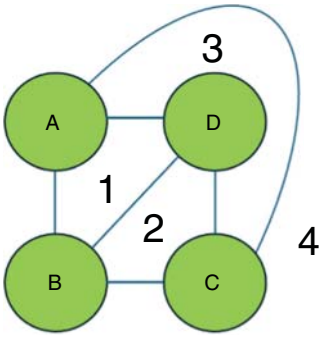
When a planar graph is drawn without edges crossing, the edges and vertices of the graph divide the plane into regions. We will call each region a **face**. Let us make this notion explicit by offering a visual example.



Recall that a planar graph is a graph that can be drawn on a plane (or a two-dimensional surface) without any edges crossing each other. In this illustration, it appears edges do cross one another, but this is only due to the presentation we have selected. The edge connecting A to C (since its positioning is arbitrary) could be changed to:



In doing so, we see the graph subdivides the plane into a number of faces. We number them, stating as we do so that the numbering is arbitrary and only intended to show that the faces are distinct.



The face labeled "4" encompasses all of the plane NOT enclosed by faces 1, 2, and 3.

Now, let us turn to Euler's formula; it states in a connected planar graph:

$$V - E + F = 2$$

Here's how Euler's formula works:

- V represents the number of vertices in the planar graph.
- E represents the number of edges in the planar graph.
- F represents the number of faces in the planar graph, including the outer (unbounded) face.

Let us apply Euler's formula to the planar graph example from above:

There are four vertices (A, B, C, D), six edges (AB, AC, AD, BC, BD, CD), and (as mentioned a moment ago) four faces. Euler's formula tells us $4 - 6 + 4$ should equal 2—which it does.

In this graph:

- V = 5 (vertices A, B, C, D, E).
- E = 6 (edges {A, B}, {A, C}, {A, D}, {B, E}, {C, D}, {D, E}).
- F = 3 (faces: the outer face, face {A, B, D}, and face {A, C, D}).

Applying Euler's formula:

$V - E + F = 5 - 6 + 3 = 2$

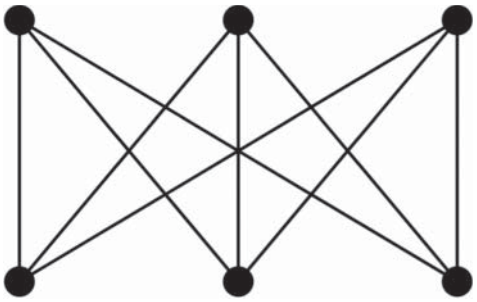
The formula holds true for this planar graph.

Euler's formula provides a necessary condition for a graph to be planar. If a connected graph satisfies Euler's formula, it is a candidate for being planar. However, satisfying Euler's formula alone is not sufficient to guarantee planarity. There are additional conditions and theorems, such as Kuratowski's theorem and Wagner's theorem, which provide more comprehensive characterizations of planar graphs.

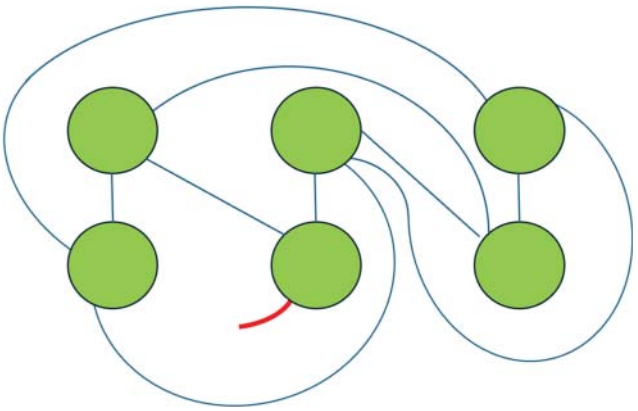
Euler's formula has several implications and applications:

- It gives an upper bound on the number of edges in a planar graph. For a connected planar graph with V vertices ($V \geq 3$), the maximum number of edges is $3V - 6$.
- It helps in determining the minimum number of colors required to color the faces of a planar graph such that no two adjacent faces have the same color (the famous four color theorem).
- It is used in algorithms for testing graph planarity and finding planar embeddings of graphs.

Example 12.39 Is the following graph planar?



Solution:
If the graph is planar, then it satisfies Euler's formula. We notice immediately that V = 6, E = 9. The number of faces F is not clear, initially, so we redraw the graph so there are no intersections of the edges.



Note the edge drawn in red, emanating from the middle vertex in the lower row. For the graph to be connected, it must be able to be joined to the vertex in the upper right, but this cannot be done without crossing another edge. The graph, therefore, is not planar.

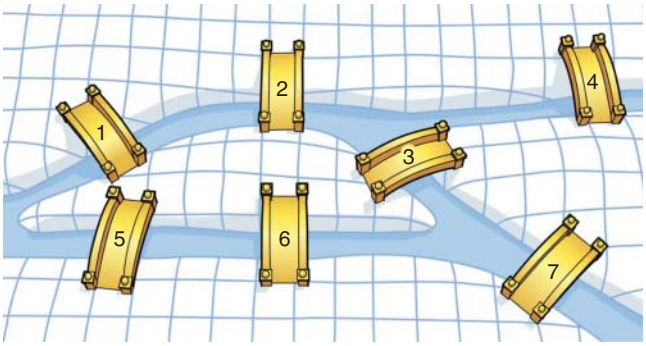
This example is an instantiation of the "utilities problem." If the three upper-row vertices represent houses, and the three lower-row vertices represent utilities (gas, water, electricity), it is not possible to connect the three houses to the three utilities without crossing lines.

Planar graphs and Euler's formula have significant theoretical and practical importance in graph theory and its applications. Understanding the properties and constraints of planar graphs is crucial for solving problems related to graph embedding, graph coloring, and network design.

12.7.3 Eulerian Graphs

Eulerian graphs are a special class of graphs named again after Euler, who studied them in the context of the famous seven bridges of Königsberg problem.

The Königsberg Bridges Problem: The river Pregel divides Königsberg into four separate parts, which are connected by seven bridges. Is it possible to walk around the city crossing all of the bridges exactly once—but not more than once? (You can start and finish anywhere, not necessarily in the same place.)

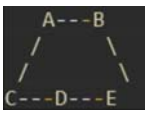


An Eulerian graph is a connected graph that contains an Eulerian circuit, which is a closed walk that visits every edge exactly once.

12.7.3.1 Definition

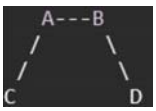
An **Eulerian Circuit** in a graph is a closed walk that traverses every edge of the graph exactly once, returning to its initial vertex following those traversals. A graph is called **Eulerian** if it contains an Eulerian Circuit. A graph is called **semi-Eulerian** if there is a path that starts and finishes at different vertices and that uses every edge.

Here's an example of an Eulerian graph:



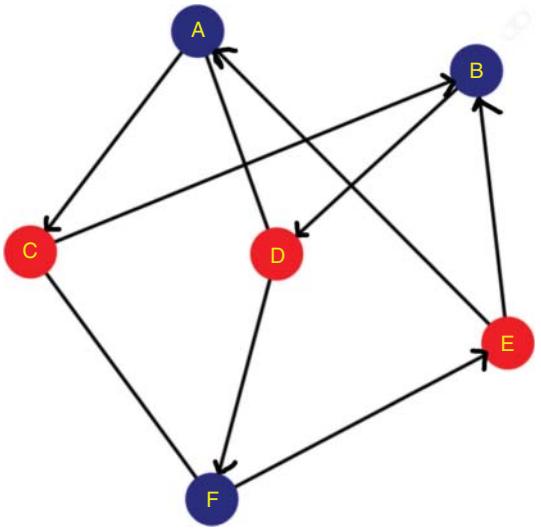
In this graph, an Eulerian circuit can be found: A – B – E – D – C – A. This circuit visits every edge exactly once and returns to the starting vertex.

On the other hand, here's an example of a non-Eulerian graph:



In this graph, it is impossible to find a closed walk that visits every edge exactly once. The vertex D has an odd degree (1), making it impossible to have an Eulerian circuit.

Example 12.40 Is this graph Eulerian?



Solution:
Clearly, this is a directed graph. The question is: can we visit every vertex while traversing each edge exactly once?

If we start at vertex A, we can proceed as A–C–B–D–F–E–A. Is this an Eulerian Circuit? No; the edge connecting F to C is not traversed on this path. How do we KNOW there is not another path that DOES form an Eulerian Circuit? The answer comes to us next.

12.7.3.2 Necessary and Sufficient Conditions

For a graph to be Eulerian, it must satisfy the following necessary and sufficient conditions:

- All vertices in the graph must have even degree.
- The graph is connected.

For a graph to be semi-Eulerian, it must satisfy the following necessary and sufficient conditions:

- There are exactly two vertices of odd degree.
- The graph is connected.

Let us illustrate these conditions with an example:

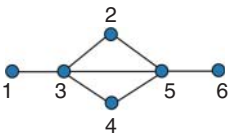


In this graph:

1) Vertex A has degree 2.
2) Vertex B has degree 2.
3) Vertex C has degree 2.
4) Vertex D has degree 2.
5) Vertex E has degree 3.
6) Vertex F has degree 1.

There are two vertices with odd degree, and the graph is connected. Therefore, this graph is semi-Eulerian.

Example 12.41 Classify the following graph as Eulerian, semi-Eulerian, or neither.



Solution:
There are two vertices with odd degree (1 and 6), and all other vertices have even degree. This graph is semi-Eulerian. A traversal of the graph that passes across each of the edges exactly once is 1 – 3 – 2 – 5 – 3 – 4 – 5 – 6.

12.7.3.3 Eulerian and Semi-Eulerian Trails

If a graph has a closed trail (it starts and finishes at the same vertex) that uses every edge, it is called Eulerian, and any such trail is called an Eulerian trail or an Eulerian circuit.

If a graph has an open trail (it starts and finishes at different vertices) that uses every edge, it is described as semi-Eulerian. Such a trail (which must be unique) is called a semi-Eulerian trail.

Eulerian graphs and trails have various applications, including:

- Solving puzzles and problems related to traversing all edges of a graph, such as the seven bridges of Königsberg problem.
- Finding optimal routes or tours in network design and logistics, where every edge needs to be visited exactly once.
- Sequencing and scheduling problems, where tasks or operations need to be performed in a specific order without repetition.

12.7.4 Hamiltonian Graphs

When we looked at Eulerian graphs, we were focused on using each of the **edges** just once. We now consider the case where **vertices** are visited exactly once.

Hamiltonian graphs, named after the Irish mathematician Sir William Rowan Hamilton, are graphs that contain a Hamiltonian cycle, which is a closed path that visits every vertex exactly once, except for the starting and ending vertex, which is visited twice (because, being a cycle, the starting and ending vertices coincide).

A Hamiltonian graph can contain numerous Hamiltonian cycles.

Here is an example of a Hamiltonian graph:



In this graph, a Hamiltonian cycle can be found: A–B–E–D–C–A. This cycle visits every vertex exactly once and returns to the starting vertex.

On the other hand, here's an example of a non-Hamiltonian graph:



In this graph, it is impossible to find a closed path that visits every vertex exactly once. There is no way to include vertex D in a cycle that visits every other vertex exactly once.

If there is a traversal of the graph that visits all the vertices exactly once but does NOT return to the starting vertex, we call it a semi-Hamiltonian graph.

In a complete graph having n vertices, (by definition) each vertex is connected to every other vertex by an edge.

Starting at any vertex, there are $n-1$ possible choices of vertices to visit next. After that, there are $n-2$ vertices left to visit, and so on. Since every vertex is connected to every other vertex, this means that there are $n! = n \times (n-1) \times (n-2) \times \times \times \times \times 2 \times 1$ possible Hamiltonian cycles in a complete graph.

12.7.4.1 Necessary Conditions

Unlike Eulerian graphs, there are no known necessary and sufficient conditions for a graph to be Hamiltonian. However, there are several necessary conditions that a Hamiltonian graph must satisfy:

1) The graph must be connected.
2) The degree of each vertex must be at least 2.
3) If the graph has n vertices ($n \geq 3$), then for every non-empty proper subset S of the vertices, the number of components in the subgraph induced by the vertices not in S must be less than or equal to |S|.

These conditions are necessary but not sufficient, meaning that a graph satisfying these conditions may or may not be Hamiltonian.

Here's an example of a graph that satisfies the necessary conditions but is not Hamiltonian:



In this graph:

1) The graph is connected.
2) All vertices have a degree of at least 2.
3) For any non-empty proper subset S of vertices, the number of components in the subgraph induced by the vertices not in S is less than or equal to |S|.

However, this graph does not have a Hamiltonian cycle. It is impossible to find a closed path that visits every vertex exactly once.

12.7.4.2 Sufficient Conditions

While there are no known necessary and sufficient conditions for a graph to be Hamiltonian, there are several sufficient conditions that guarantee the existence of a Hamiltonian cycle in a graph. Two well-known sufficient conditions are:

- Dirac's theorem: If a simple graph with n vertices ($n \geq 3$) has a minimum degree of at least $n/2$, then the graph is Hamiltonian.
- Ore's theorem: If a simple graph with n vertices ($n \geq 3$) satisfies the condition that the sum of the degrees of any two nonadjacent vertices is at least n, then the graph is Hamiltonian.

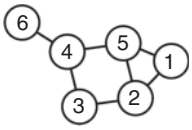
These sufficient conditions provide a way to determine if a graph is definitely Hamiltonian, but a graph not satisfying these conditions may still be Hamiltonian.

Hamiltonian graphs and cycles have various applications, including:

1) Solving the traveling salesman problem (TSP), where the objective is to find the shortest possible route that visits each vertex exactly once and returns to the starting vertex.
2) Designing efficient algorithms for problems related to graph traversal and optimization, such as finding the shortest Hamiltonian path or cycle.
3) Modeling and solving problems in scheduling, logistics, and network design, where a specific order of visiting vertices is required.

Determining whether a graph is Hamiltonian is a challenging problem, and there is no known efficient algorithm for solving it in the general case. One can attempt to establish these properties by a brute-force technique, producing the Hamiltonian or semi-Hamiltonian trails, but for large graphs, this could be a daunting process. The Hamiltonian cycle problem is NP-complete, meaning that finding a Hamiltonian cycle in a graph is computationally difficult for large graphs.

Example 12.42 Is the shown graph Hamiltonian, semi-Hamiltonian, or neither?



Solution:

Clearly, the graph is not Hamiltonian, as vertex 6 (once visited) cannot be passed without revisiting 4, and if vertex 6 is the starting vertex, it is impossible to find a path of return that does not pass through vertex 4 twice.

Is it semi-Hamiltonian? Starting at vertex 6, we perform the following traversal: 6–4–5–1–2–3.

As an aside, is the graph Eulerian? Is it semi-Eulerian? Why?

12.8 Graph Algorithms

12.8.1 Shortest Path Algorithms

Shortest path algorithms aim to find the path with the minimum total weight or distance between two vertices in a weighted graph. These algorithms have wide applications in network routing, navigation systems, and optimization problems.

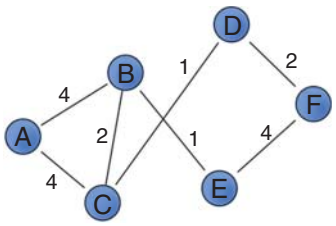
12.8.1.1 Dijkstra's Algorithm

Dijkstra's algorithm (named after Edsger W. Dijkstra [1930–2002]) is a single-source shortest path algorithm that finds the shortest path from a source vertex to all other vertices in a weighted graph with non-negative edge weights. It works by maintaining a set of visited vertices and iteratively selecting the vertex with the minimum distance from the source.

The algorithm:

1) Set initial distances for all vertices: 0 for the source vertex, and infinity for all the other.
2) Choose the unvisited vertex with the shortest distance from the start to be the current vertex. So, the algorithm will always start with the source as the current vertex.
3) For each of the current vertex's unvisited neighbor vertices, calculate the distance from the source and update the distance if the new, calculated distance is lower.
4) We are now done with the current vertex, so we mark it as visited. A visited vertex is not checked again.
5) Return to step 2, choose a new current vertex, and keep repeating these steps until all vertices are visited.
6) In the end, we are left with the shortest path from the source vertex to every other vertex in the graph.

Here's a step-by-step example of Dijkstra's algorithm:
Consider the following weighted graph:



Let's find the shortest path from vertex A to all other vertices.
Initialize the distance of vertex A to 0 and all other vertices to infinity.

- Distance[A] = 0.
- Distance[B] = ∞.
- Distance[C] = ∞.
- Distance[D] = ∞.
- Distance[E] = ∞.
- Distance[F] = ∞.

There is a tie for the nearest vertex to A, as both vertex B and vertex C are 4 units from vertex A. We will enter those values (possibly temporarily) as the distance from A to the respective vertices.

- Distance[A] = 0.
- Distance[B] = 4.
- Distance[C] = 4.
- Distance[D] = ∞.
- Distance[E] = ∞.
- Distance[F] = ∞.

C and E are adjacent to B; along the path from vertex A passing through B, the distances from A to the respective vertices are d[C] = 6 (that is higher than the already-registered distance to C, so we disregard this result), and d[E] = 4 + 1 = 5. We update our list of distances.

- Distance[A] = 0.
- Distance[B] = 4.
- Distance[C] = 4.
- Distance[D] = ∞.
- Distance[E] = 5.
- Distance[F] = ∞.

B and D are adjacent to C; along the path from vertex A passing through C, the distances from A to the respective vertices are d[B] = 6 (that is higher than the already-registered distance to B, so we disregard this result), and d[D] = 4 + 1 = 5. We update our list of distances.

- Distance[A] = 0.
- Distance[B] = 4.
- Distance[C] = 4.
- Distance[D] = 5.
- Distance[E] = 5.
- Distance[F] = ∞.

The remaining vertex, F, is adjacent to both D and E; traversing from A to F through D gives a distance of 7, while moving from A to F through E gives us a distance of 9. Therefore, the distance from A to F must be 7, and we update our distances:

- Distance[A] = 0.
- Distance[B] = 4.
- Distance[C] = 4.
- Distance[D] = 5.
- Distance[E] = 5.
- Distance[F] = 7.

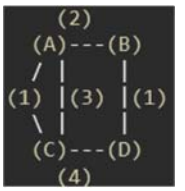
Dijkstra's algorithm has a time complexity of O((V + E) log V) when implemented using a priority queue, where V is the number of vertices, and E is the number of edges in the graph.

12.8.1.2 Floyd–Warshall Algorithm

The Floyd–Warshall algorithm (named after its creators Robert Floyd and Stephen Warshall) is an all-pairs shortest path algorithm that finds the shortest path between every pair of vertices in a weighted graph. It works by iteratively updating the distances between all pairs of vertices based on the intermediate vertices.

Here's a step-by-step example of the Floyd–Warshall algorithm:

Consider the following weighted graph:



Let us find the shortest path between every pair of vertices.

- Initialize a distance matrix dist with the direct distances between vertices. If there is no direct edge, set the distance to infinity.

	A	B	C	D
A	0	2	1	∞
B	2	0	3	1
C	1	3	0	4
D	∞	1	4	0

- Update the distance matrix by considering each vertex as an intermediate vertex. For each pair of vertices (i, j), if the path through an intermediate vertex k is shorter than the current direct distance, update the distance.

Iteration 1 (Intermediate vertex: A):

	A	B	C	D
A	0	2	1	∞
B	2	0	3	1
C	1	3	0	4
D	∞	1	4	0

Iteration 2 (Intermediate vertex: B):

	A	B	C	D
A	0	2	1	3
B	2	0	3	1
C	1	3	0	4
D	3	0	4	0

Iteration 3 (Intermediate vertex: C):

	A	B	C	D
A	0	2	1	3
B	2	0	3	1
C	1	3	0	4
D	3	1	4	0

Iteration 4 (Intermediate vertex: D):

	A	B	C	D
A	0	2	1	3
B	2	0	3	1
C	1	3	0	4
D	3	1	4	0

The final distance matrix represents the shortest paths between every pair of vertices:

	A	B	C	D
A	0	2	1	3
B	2	0	3	1
C	1	3	0	4
D	3	1	4	0

The Floyd–Warshall algorithm has a time complexity of $O(V^3)$, where V is the number of vertices in the graph. It is useful for finding shortest paths in dense graphs or when multiple shortest path queries need to be answered efficiently.

By way of comparison, Dijkstra's algorithm is efficient for single-source shortest paths in sparse graphs, while the Floyd–Warshall algorithm is suitable for all-pairs shortest paths in dense graphs or when multiple queries are required.

12.8.2 Minimum Spanning Tree Algorithms

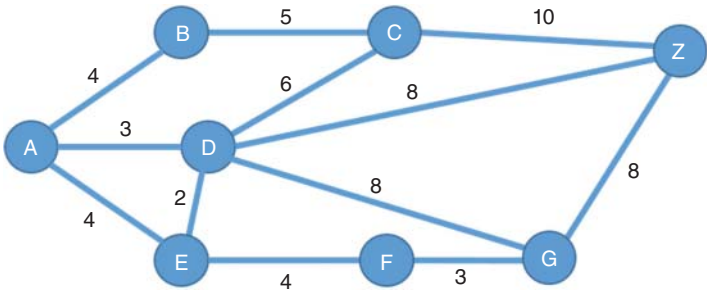
An MST is a tree that spans all the vertices of a weighted, connected graph while minimizing the total weight of the edges. In this section, we will discuss two popular MST algorithms: Kruskal's algorithm and Prim's algorithm.

12.8.2.1 Kruskal's Algorithm

Kruskal's algorithm (Joseph Bernard Kruskal, Jr., 1928–2010) is a greedy algorithm (by this we mean it makes a locally optimal choice in each step in order to find the optimal solution) that finds the MST of a weighted, connected graph. It works by sorting the edges in nondecreasing order of their weights and iteratively adding edges to the MST while avoiding cycles.

Here's a step-by-step example of Kruskal's algorithm:

Consider the following weighted graph:



Let us find the MST using Kruskal's algorithm.

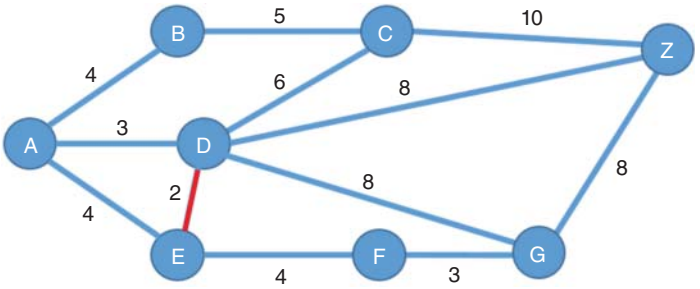
1) Sort the edges in nondecreasing order of their weights:

Edge	Weight
(D,E)	2
(F,G)	3
(A,D)	3
(A,B)	4
(A,E)	4
(E,F)	4
(B,C)	5
(C,D)	6
(D,Z)	8
(D,G)	8
(G,Z)	8
(C,Z)	10

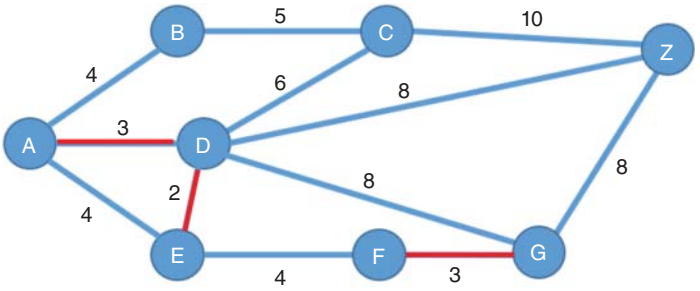
Begin with the smallest weight; if it does not form a cycle with the weights found so far, discard it. Otherwise, include this edge.

As we proceed, we will highlight the "kept" edges.

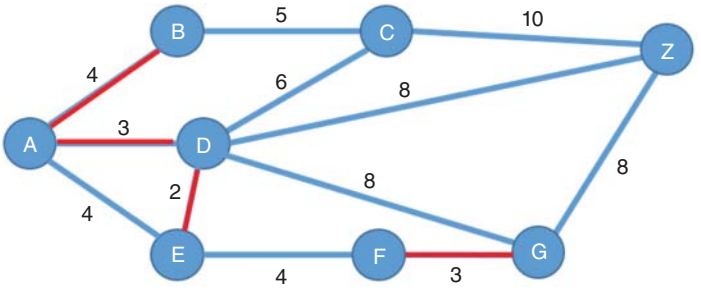
Thus, we start with (D, E).



Next, we consider edges (F, G) and (A, D). Neither of those forms a cycle, so we keep both:

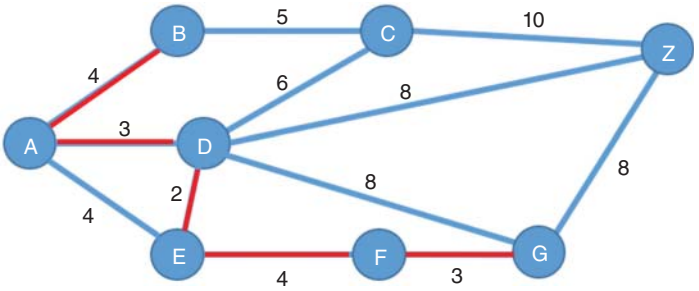


Next, notice that edge (A, B) does not create a cycle, so we include it:

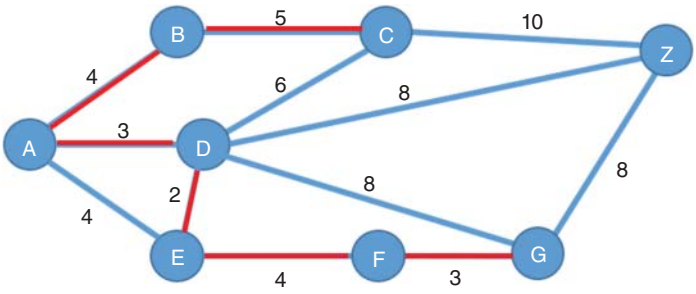


The next edge to consider, (A, E), would form a cycle, so we discard it.

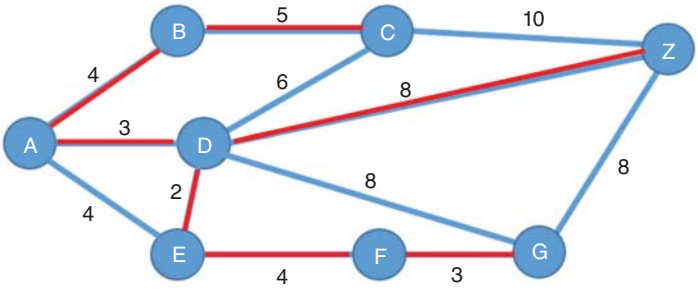
The next edge in the table is (E, F). It does not form a cycle, so we keep it.



The next edge is (B, C); it does not form a cycle, so we keep it:



Edge (C, D) would form a cycle, so we discard it. The next edge, (D, Z), does not form a cycle, so we keep it:



Edge (G, Z) would form a cycle, so we discard it. Similarly, edge (C, Z) would form a cycle, so again we discard.

The MST is the graph shown, omitting all edges that remain blue. The total spanning weight of the graph is 29.

Kruskal's algorithm has a time complexity of O(E log E) or O(E log V), where E is the number of edges, and V is the number of vertices in the graph. It is efficient for sparse graphs and can be implemented using a disjoint-set data structure to efficiently detect cycles.

12.8.2.2 Prim's Algorithm

Prim's algorithm, established in 1930 by Vojtěch Jarník and rediscovered and republished by computer scientists Robert C. Prim in 1957, is another greedy algorithm for finding the MST of a weighted, connected graph. It starts with an arbitrary vertex and iteratively grows the MST by adding the nearest vertex that is not yet included in the MST.

Here are the steps we must follow:

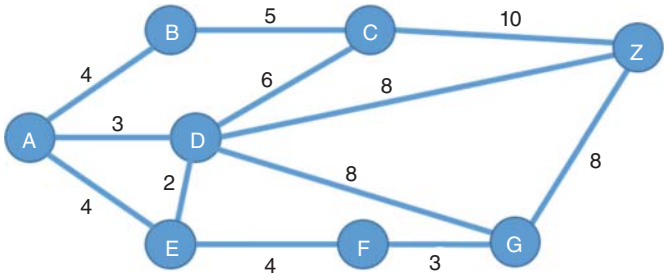
Step 1: *Determine an arbitrary vertex as the starting vertex of the MST.*

Step 2: *Follow steps 3 to 5 till there are vertices that are not included in the MST (known as* ***fringe*** *vertices).*

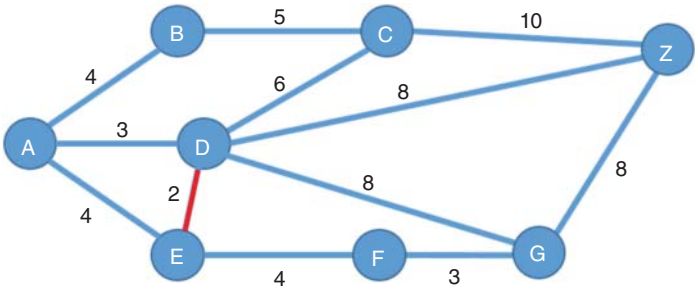
Step 3: *Find edges connecting any tree vertex with the fringe vertices.*

Step 4: *Find the minimum among these edges.*
Step 5: *Add the chosen edge to the MST if it does not form any cycle.*
Step 6: *Return the MST and exit*

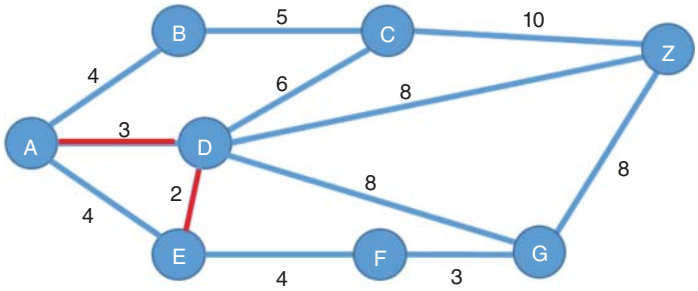
Here's a step-by-step example of Prim's algorithm:
Consider the same weighted graph as before:
Let us find the MST starting from vertex A.



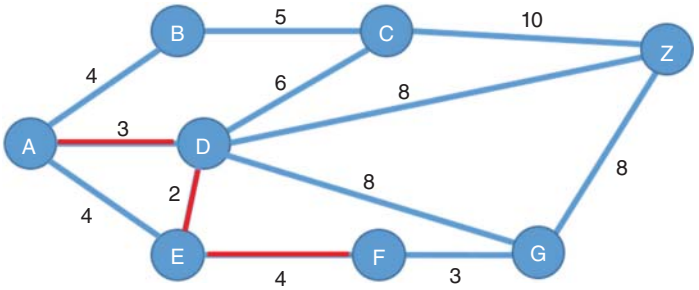
We will arbitrarily begin using the vertex E; the fringe vertices from E are A, D, and F. We compare the weights of the edges connecting E to these vertices and choose the smallest weight; therefore, we will use edge (E, D).



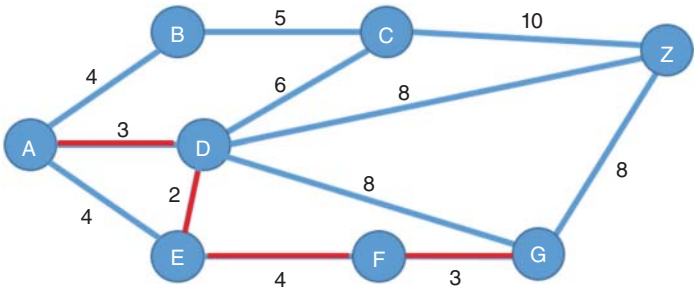
The tree now contains vertices D and E; the fringe vertices from these are A, C, Z, G, and F. From among the weights connecting either D or E to any fringe vertex, we choose the smallest weight. That would be edge AD.



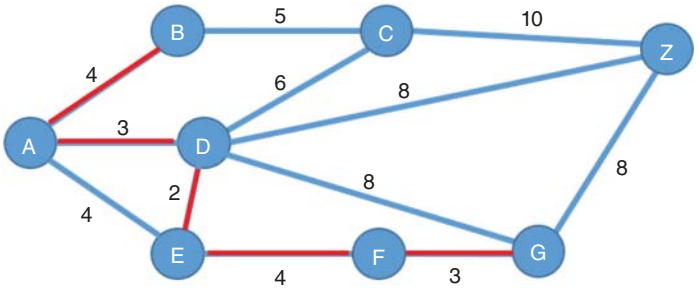
The fringe vertices are now B, C, Z, F, and G. The smallest weight connecting a vertex in the tree to one of these fringes is 4; we include edge (E, F) in the tree:



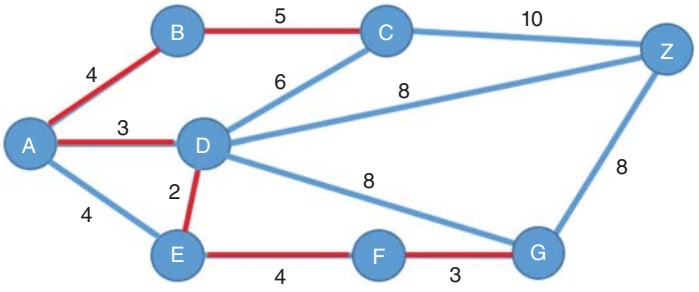
The fringe vertices are now B, C, Z, and G.
The smallest weight from the tree to a fringe vertex is 3, and so we include edge (F, G) in the tree:



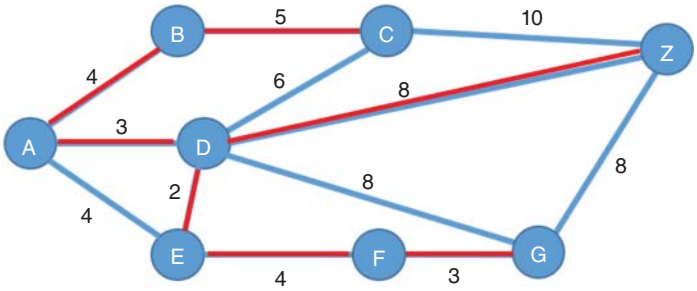
The fringe vertices are now B, C, and Z. The smallest weight connecting the tree to a fringe vertex is 4, so we include edge (A, B).



The fringe vertices are now C and Z. The smallest weight connecting the tree to a fringe vertex is 5, so we add edge (B, C) to the tree:



The only remaining fringe vertex is Z. The smallest weight connecting the tree to Z has weight 8, so we use one of the two choices of edge that has that weight; we will use (D, Z).



We now have a spanning tree, consisting of the red edges. Observe that the total weight of the graph is 29, the same as before.

Prim's algorithm has a time complexity of O((V + E) log V) when implemented using a priority queue, where V is the number of vertices, and E is the number of edges in the graph. It is efficient for dense graphs and can be implemented using an adjacency matrix or an adjacency list representation of the graph.

Both Kruskal's algorithm and Prim's algorithm guarantee to find the MST of a weighted, connected graph. They have different approaches but yield the same result. Kruskal's algorithm is based on sorting edges, while Prim's algorithm grows the MST from a starting vertex.

MST algorithms have various applications, including network design, clustering, and approximation algorithms for NP-hard problems. They help in finding the most cost-effective way to connect all vertices in a graph while minimizing the total weight of the edges.

12.8.3 Network Flow Algorithms

Network flow algorithms address the problem of finding the maximum flow through a network, which is a directed graph with a source vertex and a sink vertex, and each edge has a capacity representing the maximum amount of flow that can pass through it. The goal is to find the maximum flow from the source to the sink while respecting the capacity constraints of the edges. In this section, we will discuss the Ford–Fulkerson algorithm and the Max-Flow Min-Cut theorem.

12.8.3.1 Ford–Fulkerson Algorithm

The Ford–Fulkerson algorithm is a greedy algorithm that finds the maximum flow in a network. It works by repeatedly finding augmenting paths from the source to the sink and updating the flow along these paths until no more augmenting paths can be found. An **augmenting path** refers to a path in a flow network that can accommodate more flow, thereby increasing the total flow value from the source to the sink.

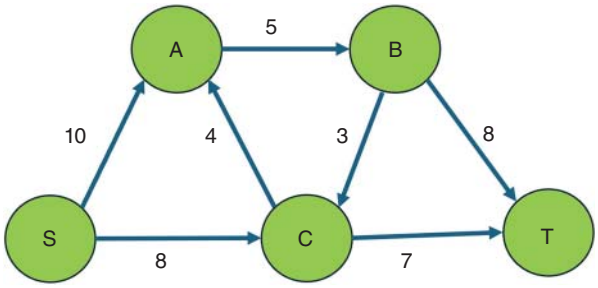
Note: The Ford–Fulkerson algorithm is often described as a *method* instead of as an *algorithm*, because it does not specify how to find a path where the flow can be increased. This means it can be implemented in different ways, resulting in different time complexities.

The steps in the algorithm/method are:

- Start with zero flow on all edges.
- Find an *augmented path* where more flow can be sent.
- Do a *bottleneck calculation* to find out how much flow can be sent through that augmented path.

- Increase the flow found from the bottleneck calculation for each edge in the augmented path.
- Repeat steps 2–4 until max flow is found. This happens when a new augmented path can no longer be found.

Here's a step-by-step example of the Ford–Fulkerson algorithm:
Consider the following network:



There is no flow in the graph initially.

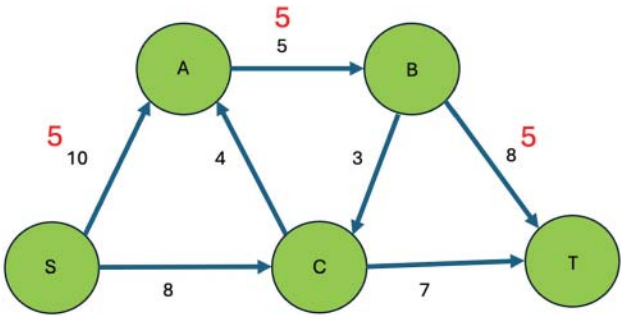
In order to find the maximum flow, we need to formalize the notion of a "bottleneck" in the network. Intuitively, the flow cannot be increased forever because there is some subset of edges whose capacities eventually become saturated with flow.

Every path from s to t must cross one of these saturated edges, and so the sum of capacities of these edges imposes an upper bound on the size of the maximum flow. Thus, these edges form a bottleneck.

Let us find the maximum flow from the source vertex S to the sink vertex T.

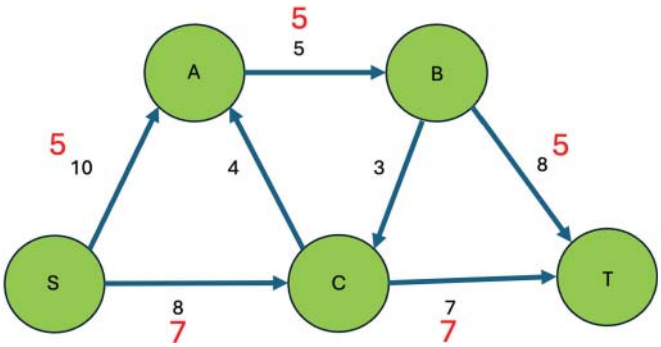
1) Initialize the flow in each edge to 0.
2) While there exists an augmenting path from S to T:
   - Find an augmenting path using DFS or BFS.
   - Compute the minimum residual capacity along the path.
   - Update the flow along the path by adding the minimum residual capacity to the forward edges and subtracting it from the backward edges.

The first augmenting path we can find is S–A–B–T; the minimum of the flow rates along this path is 5, and therefore, we can increase the flow along this path by 5.

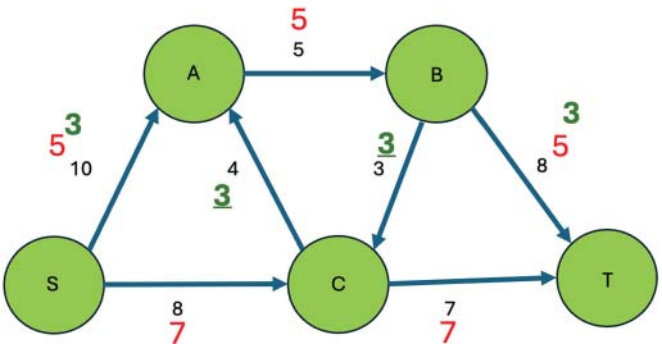


The original weights show the available flow through each edge, and the red numbers show the flow actually going through the edges (because there is a capacity limitation at some point along the flow path).

Another augmenting path we can find is S–C–T;the minimum flow capacity along that path is 7, so we can increase the flow in that direction by 7; the total flow through the network is now $5 + 7 = 12$.



Another augmenting path we can find is $S - A - C - B - T$; the residual flows along this path are 5, 4, 3, and 3. The minimum residual flow is 3, and so we can increase the flow in that direction to 3. Note, however, we are going against some of the directional arrows on the graph! This means the overall flow along this path will be (shown in green):



The flow along this path is 3, which brings the total flow through the network to $12 + 3 = 15$.

There are no other augmenting paths possible; though there is still a residual capacity flow of 1 unit from S to C and from C to A, the remaining paths to the sink T are saturated, and no more flow is possible.

The Ford–Fulkerson algorithm has a time complexity of O(E * max_flow), where E is the number of edges in the network, and max_flow is the maximum flow value. The efficiency of the algorithm depends on the choice of the augmenting path-finding method.

Using BFS to find the shortest augmenting path results in the Edmonds–Karp algorithm, which has a better time complexity of $O(V * E^2)$, where V is the number of vertices.

12.8.3.2 Max-Flow Min-Cut Theorem

The Max-Flow Min-Cut theorem states that the maximum flow in a network is equal to the minimum capacity of all s–t cuts in the network, where an s–t cut is a partition of the vertices into two disjoint sets, one containing the source vertex s and the other containing the sink vertex t.

In other words, the maximum amount of flow that can be pushed from the source to the sink is limited by the minimum capacity of the edges that need to be removed to disconnect the source from the sink.

The Max-Flow Min-Cut theorem provides a way to verify the correctness of the maximum flow found by algorithms like Ford–Fulkerson. It also has practical applications in network design, where the minimum cut represents the bottleneck or the weakest link in the network.

Finding the maximum flow and the minimum cut in a network has various real-world applications, such as:

1) Transportation networks: Maximizing the flow of vehicles or goods through a network of roads or channels.
2) Communication networks: Optimizing the flow of data or information through a network of routers or switches.
3) Bipartite matching: Finding the maximum matching in a bipartite graph, which can be formulated as a network flow problem.
4) Image segmentation: Segmenting an image into foreground and background regions based on the minimum cut of a graph representing pixel similarities.

13

What Is Cryptography?

13.1 What Is Cryptography?

Cryptography is a fundamental concept in the field of cybersecurity, with applications spanning various domains, including computer science, information security, and communication systems.

13.1.1 Definition and Purpose

Cryptography is the study and practice of techniques for secure communication in the presence of adversaries. The primary goal of cryptography is to enable two or more parties to communicate sensitive information over an insecure channel, such as the internet, without unauthorized entities being able to intercept, understand, or tamper with the transmitted data.

It is impractical to suppose communications can be completely secure. Numerous obstacles exist to prevent this: treachery, technical limitations, superior technology in possession of the opposition, and others. Encryption (the disguise of sensitive information) is paramount to the transmission of information in such a manner as to make communication as secure as possible.

The process of transforming native information (data "in the clear") into coded information is called "encryption." The inverse of this process, the transformation of coded information into native information, is called "decryption."

The core principles of cryptography revolve around the following key concepts:

1) Confidentiality: Ensuring that the information remains private and accessible only to authorized parties.
2) Integrity: Protecting the data from unauthorized modification or tampering.
3) Authentication: Verifying the identity of the communicating parties to prevent impersonation.
4) Non-repudiation: Preventing any party from denying their involvement in a communication or transaction.

Cryptographic techniques employ mathematical algorithms and protocols to achieve these four goals. The two main branches of cryptography are:

1) Symmetric-key cryptography: Both the sender and the receiver use the same secret key for encryption and decryption.
2) Asymmetric-key (or public-key) cryptography: Each party has a pair of keys—a public key for encryption and a private key for decryption.

Cryptography finds extensive use in various applications, such as secure web browsing (e.g., Hypertext Transfer Protocol Secure (HTTPS)), email encryption (e.g., Pretty Good Privacy (PGP)),

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

digital signatures, transmission of military intelligence and classified information, and cryptocurrency transactions (e.g., Bitcoin).

13.1.2 Brief History of Cryptography

The history of cryptography dates back to ancient times when secret writing and hidden messages were used for military and diplomatic purposes. Some notable cryptographic examples include:

1) Caesar Cipher (100 BCE–44 BCE): A simple substitution cipher used by Julius Caesar to communicate with his generals.
2) Al-Kindi's Book on Cryptanalysis (9th century CE): The first known recorded explanation of cryptanalysis.
3) Vigenère Cipher (16th century): A polyalphabetic substitution cipher that remained unbroken for centuries.
4) Enigma Machine (early 20th century): An electromechanical rotor cipher machine used by Nazi Germany during World War II.

The advent of computers and the digital age brought about a revolution in cryptography. In the 1970s, the Data Encryption Standard (DES) was introduced as a symmetric key algorithm for commercial use. However, the rapid increase in computational power and development of programming techniques rendered DES vulnerable, leading to the development of Advanced Encryption Standard (AES) in 2001.

Public-key cryptography, also called asymmetric cryptography, introduced by Diffie and Hellman in their 1976 paper, *New Directions in Cryptography*, marked a significant milestone in the field. The Rivest-Shamir-Adleman (RSA) algorithm, developed by (and named using the first letter of the surnames of) Rivest, Shamir, and Adleman in 1978, became the most widely used public-key cryptosystem.

> *Note: In actuality, British mathematician Clifford Cocks developed an equivalent encryption system to RSA, five years earlier, in his work for the Government Communications Headquarters (GCHQ). This was classified (under the Official Secrets Act) and used by the Signals Intelligence Agency (SIGINT), but declassified in 1997 by the British government.*

Today, cryptography continues to evolve, with ongoing research in areas such as elliptic curve cryptography (ECC), quantum cryptography, and homomorphic encryption, aiming to address ever-growing challenges in secure communication and data protection.

The importance of cryptography cannot be overstated, nor can the use of human ingenuity as an adjunct to computational processes. During the Second World War, the Japanese navy employed an effective encryption methodology that was broken by Allied intelligence, yet remained secure due to the use of alternative identifiers. The military installations on the island of Midway were code-named AF by the Japanese, but this identification was only determined when deliberate misinformation was released to the Japanese that caused them to inadvertently reveal that sobriquet.

13.2 Basic Cryptographic Concepts

To understand the workings of cryptography, it is essential to grasp the fundamental concepts and terminology used in the field. This section introduces the basic cryptographic concepts, including

plaintext, ciphertext, keys, encryption, decryption, and the distinction between symmetric and asymmetric cryptography.

13.2.1 Plaintext, Ciphertext, and Keys

In cryptography, the original message or data that needs to be protected is called the **plaintext**. It is the unencrypted, readable form of the information. The plaintext can be any type of data, such as text, numbers, images, or audio files. When transmitted in this form, we say the data has been sent "in the clear." The risk of exposure of such information is high.

To secure the plaintext, it is transformed into an unintelligible form (or, at the very least, hopefully unintelligible form) called the **ciphertext**. The ciphertext is the encrypted version of the message that appears as a random sequence of characters or bits.

With more formality than already stated, we define the process of converting plaintext into ciphertext as **encryption**.

13.2.2 Encryption, Decryption, and Hacking

Encryption is the process of converting plaintext into ciphertext using a cryptographic algorithm and a key. The purpose of encryption is to protect the confidentiality of the message by making it unreadable to anyone who does not possess the correct key.

Decryption is the reverse process of encryption, wherein the ciphertext is transformed back to plaintext using the same cryptographic algorithm and the corresponding key. Only intended recipients possessing the correct key can decrypt the ciphertext and retrieve the original message.

Hacking is the act of compromising digital devices and networks through unauthorized access to an account or computer system. While not always malicious, it is most associated with illegal activity and data theft.

In the context of cybersecurity, hacking refers to the misuse of devices like computers, smartphones, tablets, and networks to cause damage to or corrupt systems, gather information on users, steal data and documents, or disrupt data-related activity.

Mathematically, the encryption and decryption processes can be represented as follows:

- Encryption: $C = E(P, K)$, where C is the ciphertext, P is the plaintext, K is the key, and E is the encryption function.
- Decryption: $P = D(C, K)$, where P is the plaintext, C is the ciphertext, K is the key, and D is the decryption function.

13.2.3 Symmetric and Asymmetric Cryptography

Cryptographic algorithms can be classified into two main categories: symmetric and asymmetric.

Symmetric cryptography, also known as secret-key cryptography, uses the same key for both encryption and decryption. Sender and the recipient must share the secret key beforehand through a secure channel or some preexisting arrangement.

The notion of "balance" between the algorithms being used on each end of the data transmission is the source of the term "symmetric." Some common symmetric encryption algorithms include AES, DES, and ChaCha20.

There are advantages to symmetric cryptography:

- Fast and efficient for encrypting large amounts of data.
- Relatively simple to implement and use.

Predictably, disadvantages also exist:

- Requires a secure key exchange mechanism to share the secret key.
- Management of keys becomes more difficult when the number of communicating parties increases.

Asymmetric cryptography, also called public-key cryptography, uses a pair of keys: a **public key** and a **private key**. The public key is freely distributed and used for encryption, while the private key is kept secret by the owner and used for decryption.

Some well-known asymmetric encryption algorithms include RSA and ECC.

Advantages of asymmetric cryptography:

- Eliminates the need for a secure key exchange channel.
- Supports digital signatures and key exchange protocols.

Disadvantages of asymmetric cryptography:

- Computationally more intensive compared to symmetric cryptography.
- Slower for encrypting large amounts of data.

In practice, symmetric and asymmetric cryptography are often used together in **hybrid cryptosystems**, leveraging the strengths of both approaches. For example, asymmetric cryptography can be used to securely exchange a symmetric key, which is then used for efficient encryption and decryption of the actual data.

13.3 Classical Ciphers

Classical ciphers are the earliest forms of cryptographic techniques used to protect the confidentiality of messages. These ciphers rely on simple mathematical operations and are relatively easy to implement by hand. While most classical ciphers are no longer considered secure due to advancements in cryptanalysis and computing power, they provide a foundation for understanding the basic principles of cryptography. This section explores two main categories of classical ciphers: substitution ciphers and transposition ciphers.

13.3.1 Substitution Ciphers

Substitution ciphers work by replacing each character in the plaintext with another character based on a predetermined rule or mapping. A crude example is the creation of a transposition table, assigning each letter of the English language to another letter.

Such a table might begin with:

A ⇔ T
B ⇔ K
C ⇔ M
D ⇔ F
E ⇔ Z

and so on. In such a system, the word “BED” would be transformed to “KZF.”

The three main types of substitution ciphers are the Caesar cipher, monoalphabetic substitution, and polyalphabetic substitution.

13.3.1.1 Caesar Cipher

The Caesar cipher, named after Julius Caesar, is one of the simplest substitution ciphers. It operates by shifting each letter in the plaintext by a fixed number of positions in the alphabet. For example, with a shift of three, "A" becomes "D," "B" becomes "E," and so on. The mathematical representation of the Caesar cipher encryption is:

$$\mathrm{E}(x) = (x + k) \pmod{26}$$

The use of the mathematical operator "mod" may be new to you. It is a three-letter abbreviation for "**modulus**," and refers to a limitation of the number of digits used in representation values. A brief exposition on this notion is warranted, but an in-depth understanding lies in the province of abstract algebra.

When a numerical modulus is identified, all integers can be represented (with some loss of information) using a **modular representation**. Mathematical operations involving a modular representation are referred to as **modular arithmetic**.

Modular arithmetic is the branch of arithmetic mathematics related to the "mod" functionality. Basically, modular arithmetic is related to the computation of a "mod" of expressions. Expressions may have digits and computational symbols of addition, subtraction, multiplication, division, or any other operation.

Modular arithmetic, often referred to as "clock arithmetic," is a system of arithmetic for integers, where numbers wrap around upon reaching a certain value, known as the modulus. The use of the term "clock arithmetic" is a reference to the idea of how time is referenced; if it is now 10 AM and we are told a person will arrive in five hours, we do not say they will arrive at 15 AM (which is nonsensical, according to our understanding of the representation of time). We know the advance of five hours will take us, instead, to 3 PM.

The hour on the clock has "wrapped back around" to zero after reaching the maximum clock-hour value of 12.

Considering time of day broken into 24 hours, we can express the individual hours using "clock time," by restating the time values "mod 12."

Hour	Time	AM/PM
12	12	AM
1	1	AM
2	2	AM
3	3	AM
4	4	AM
5	5	AM
6	6	AM
7	7	AM
8	8	AM
9	9	AM
10	10	AM
11	11	AM
12	12	PM

Hour	Time	AM/PM
13	1	PM
14	2	PM
15	3	PM
etcetera		

Observe the stated time (the middle column) values have wrapped back around to the start, and the "new time" is found by subtracting 12 from the hour number—we say the hour number has been reduced "modulo twelve," or "mod 12."

When a number is stated and a modulus has been identified, the equivalent "modular representation" of the number is found by dividing that number by the modulus and determining the remainder. The remainder is the equivalent value to the stated number, called its modular representation.

Imagine we are beginning at midnight on a given day and looking ahead by 115 hours; what will be the reading on the clock at that time?

115 / 12 = 9, remainder 7. We say time 115 hours is equivalent to 7 (mod 12).

In $E(x) = (x + k) \pmod{26}$, x is the plaintext letter, k the shift value, and $E(x)$ the corresponding ciphertext letter. The decryption process uses the inverse operation:

$$D(y) = (y - k) \pmod{26}$$

where y is the ciphertext letter, k is the shift value, and $D(y)$ is the corresponding plaintext letter.

Example 13.1 The phrase "my cat is named jojo" is to be encrypted using the Caesar cipher with shift value 17.

What is the encrypted form of the phrase?

Solution: We construct a code key chart to visualize the equivalence relation.

Letter	Value of x (alphabet location)	Shift value k	$(x + k)$ (mod 26)
a	1	17	18
b	2	17	19
c	3	17	20
d	4	17	21
e	5	17	22
f	6	17	23
g	7	17	24
h	8	17	25
i	9	17	0
j	10	17	1
k	11	17	2

Letter	Value of x (alphabet location)	Shift value k	$(x + k)$ (mod 26)
l	12	17	3
m	13	17	4
n	14	17	5
o	15	17	6
p	16	17	7
q	17	17	8
r	18	17	9
s	19	17	10
t	20	17	11
u	21	17	12
v	22	17	13
w	23	17	14
x	24	17	15
y	25	17	16
z	26	17	17

The message "my cat is named jojo" can now be encrypted as:
4 16 20 18 11 0 10 5 18 4 22 21 1 6 1 6

Though Microsoft Excel is by no means a powerful encryption/decryption software, the use of the Power Query and VLOOKUP functions of Excel, combined with determined exploration, would allow the cipher to be quickly broken.

The Caesar cipher is easily broken using frequency analysis or brute-force attacks, as there are only 25 possible shift values (not counting the case where the shift is 0, which results in no encryption).

13.3.1.2 Monoalphabetic Substitution

Monoalphabetic substitution ciphers extend the concept of the Caesar cipher by allowing an arbitrary permutation of the alphabet for substitution. Each letter in the plaintext is replaced by a corresponding letter in the permuted alphabet. For example, "A" could be mapped to "Q," "B" to "W," and so on. This can be done in a trivial manner using software; a custom sort of a column of the letters of the alphabet will produce such a randomization.

Mathematically, a monoalphabetic substitution cipher can be represented as:

$$\mathrm{E}(x) = \pi(x)$$

where x is the plaintext letter, π the permutation function, and $\mathrm{E}(x)$ the corresponding ciphertext letter. The decryption process uses the inverse permutation:

$$\mathrm{D}(y) = \pi^{-1}(y)$$

where y is the ciphertext letter, π^{-1} the inverse permutation function, and $\mathrm{D}(y)$ the corresponding plaintext letter.

Example 13.2 The phrase "my cat is named jojo" is to be encrypted using the Monoalphabetic cipher with shift value 17. What is the encrypted form of the phrase?

Solution: We construct a code key chart to visualize the equivalence relation.

Permuted Letter	Value of x (alphabet location)	Shift value k	$(x + k)$ (mod 26)
a	1	17	18
h	2	17	19
y	3	17	20
u	4	17	21
s	5	17	22
k	6	17	23
g	7	17	24
e	8	17	25
f	9	17	0
v	10	17	1
b	11	17	2
m	12	17	3
c	13	17	4
l	14	17	5
n	15	17	6
q	16	17	7
p	17	17	8
i	18	17	9
r	19	17	10
d	20	17	11
w	21	17	12
x	22	17	13
z	23	17	14
o	24	17	15
j	25	17	16
t	26	17	17

The message "my cat is named jojo" can now be encrypted as:

1 4 18 17 9 22 6 18 3 25 11 16 15 16 15

It is somewhat tragic the notation E(x) appears elsewhere (such as in the expected value of a random variable), but context makes clear which instantiation of E(x) is intended.

While more secure than the Caesar cipher, monoalphabetic substitution ciphers are still vulnerable to **frequency analysis attacks**, as the letter frequencies in the ciphertext mirror those of the plaintext language.

Notice in the two encryptions thus far shown of the message "my cat is named jojo," we see:

16 20 18 11 0 10 5 18 4 22 21 1 6 1 6

And

3 1 4 18 17 9 22 6 18 3 25 11 16 15 16 15

We observe the pattern present in the last four values of each encrypted message; it follows that 1 and 16 must represent the same letter in plaintext, as must 6 and 15. This seemingly trivial observation serves as an initial step toward the destruction of the cipher.

British mathematician Alan Turing (1912–1954) used a notion along these lines in his work at breaking the German Enigma Machine code during the Second World War, conjecturing German fealty would compel them to include the words "Heil Hitler" in all their military communications.

13.3.1.3 Polyalphabetic Substitution

Polyalphabetic substitution ciphers, such as the Vigenère cipher (named after, though not invented by, Blaise de Vignere (1523–1596)), use multiple substitution alphabets to encrypt the plaintext. The selection of the substitution alphabet for each letter depends on a repeated keyword or phrase to match the length of the plaintext.

In the Vigenère cipher (long thought to be unbreakable), the encryption process can be represented mathematically as:

$$\mathrm{E}(x_i) = [x_i + k_i](\mathrm{mod}\ 26)$$

Where x_i is the ith plaintext letter, k_i is the ith letter of the repeated keyword, and $\mathrm{E}(x_i)$ is the corresponding ith ciphertext letter. The decryption process is given by:

$$\mathrm{D}(y_i) = [y_i - k_i](\mathrm{mod}\ 26)$$

Where y_i is the ith ciphertext letter, k_i is the ith letter of the repeated keyword, and $\mathrm{D}(y_i)$ is the corresponding ith plaintext letter.

Polyalphabetic substitution ciphers offer better resistance against frequency analysis when compared to monoalphabetic ciphers. However, they are still vulnerable to more advanced cryptanalytic techniques, such as the Kasiski examination and the Friedman test.

> *Note: Techniques such as the Kasiski examination (named for Friedrich W. Kasiski 1805–1881) are ingenious methods for attacking and destroying encryption techniques. Full examination of these lies beyond the scope of our conversation but are eminently worthy of deeper exploration.*

13.3.2 Transposition Ciphers

Transposition ciphers, also known as permutation ciphers, rearrange the order of characters in the plaintext according to a specific rule or pattern. The two main types of transposition ciphers discussed here are the rail fence cipher and the columnar transposition cipher.

13.3.2.1 Rail Fence Cipher

The rail fence cipher, also called the zigzag cipher, writes the plaintext characters in a zigzag pattern across a specified number of rows (rails). The ciphertext is then formed by reading the characters row by row.

For example, with three rails and the plaintext "HELLO WORLD", the encryption process would be:

H				O				L	
	E		L		W		R		D
		L				O			

The resulting ciphertext is "HOLELWRDLO".

The decryption process involves recreating the zigzag pattern with the ciphertext characters and then reading the plaintext in the original zigzag order.

It is efficient to arrange the message and blank spaces within the zigzag pattern in a matrix, as we will demonstrate in the following example.

Suppose a baseball team wishes to communicate internally they plan to trade outfielder Tanner to another team, effective tomorrow. They do not wish Tanner to know this, as he is planned to be in the current day's starting lineup. The message to be transmitted is: Tanner will be traded to the Braves tomorrow. We will use a rail fence cipher incorporating four rails (the number of rails has been chosen arbitrarily for this example).

The message has 37 letters, and so the matrix we construct will have four rows (one for each rail) and 37 columns. Using the method outlined above, we draw the matrix:

T					W					T					T					R					O					W					
	A			R		I			E		R			D		O				B		A				T		M				O		
		N		E				L		B				A		E				T		E				V		S				O		R		
			N						L						D						H						E						R				

Next, we use the rows of the matrix as the encrypted message:

TWTTROWARIERDOBATMONELBAETEVSORNLDHER.

The rail fence cipher provides minimal security and can be broken by guessing the number of rails and reconstructing the plaintext. Not knowing the number of rails is the only true challenge, but this can be overcome through experimentation.

Suppose Tanner suspects the organization is speaking about him, and he somehow intercepts the encrypted message displayed above. He suspects the team employs a rail fence cipher and decides to experiment, beginning with two rails, after confirming (by direct count) the message contains 37 columns.

The wily player creates a 2×37 matrix and constructs a zigzag pattern accordingly:

T	W	T	T	R	O	W	A	R	I	E	R	D	O	B	A	T	M	O																
	N	E	L	B	A	E	T	E	V	S	O	R	N	L	D	H	E	R																

Tanner now attempts to read the proposed decryption to see if sense can be made of the result:

TNWETLTBRAOEWTAERVISEORRDNOLBDAHTMERO

Immediately, he sees this is gibberish. He discards the notion two rails are involved and repeats the process for three rails, and so on. When he reconstructs the message with four rails, he discovers the portentous message and must plan his next move.

13.3.2.2 Columnar Transposition

As the name implies, this encryption technique involves columns of characters and a strategic rearrangement of those characters. As we will see, there will be a "keyword" involved that makes decryption possible, so we will observe a greater level of security than existed in the rail fence cipher.

The columnar transposition cipher writes the plaintext characters in a grid, row by row, and reads the ciphertext **column by column** based on a keyword. The columns are arranged in the order determined by the alphabetical order of the keyword letters. The technique demonstrates a certain cleverness, as we will see. In our first example, we will return to the ominous message intercepted by the ballplayer, Tanner, encrypted with this new method.

We are aware there are 37 characters in the message. Arbitrarily, we will choose to use rows of seven characters, and thus, the final row will contain $37 \pmod 7 = 2$ characters from the message, with the rest filled by agreed-upon "null letters" (we will adopt "X" as the null letter). We will use the keyword ORANGES for this message (the matrix we create should have as many columns as the keyword).

The text is arranged with the keyword at the top and the message below; recall the message to be transmitted is: Tanner will be traded to the Braves tomorrow.

O	R	A	N	G	E	S
T	A	N	N	E	R	W
I	L	L	B	E	T	R
A	D	E	D	T	O	T
H	E	B	R	A	V	E
S	T	O	M	O	R	R
0	W	X	X	X	X	X

Observe, the message is written across the rows.

We next number the entries in the first row alphabetically; note the letters in ORANGES appear alphabetically as A, E, G, N, O, R, S (had one of the letters in the keyword appear multiple times, we would have listed all multiplicities; an example of this will follow). Assigning ordering numbers to them, we see

A = 1
E = 2
G = 3
N = 4
O = 5
R = 6
S = 7

We now rearrange the columns of the message in this new ordering; that is, the column headed by "A" will be first, that headed by "E" will be second, and so on. Our transposed matrix is:

A	E	G	N	O	R	S
N	R	E	N	T	A	W
L	T	E	B	I	L	R
E	O	T	D	A	D	T
B	V	A	R	H	E	E
O	R	O	M	S	T	R
X	X	X	X	O	W	X

The encrypted message is read down the columns (ignoring the entries from the keyword row), and is:

NLEBOXRTOVRXEETAOXNBDRMXTIAHSOALDETWWRTERX

As you can see, we have achieved a remarkable scramble of the letters of the message!

How would this be decrypted?

Remember, the intended receiver is aware of the keyword, ORANGES. She knows the alphabetical order of the letters of that keyword, and thus can decode the message using the following steps:

Being aware that the keyword contains seven letters, she knows the matrix will have seven columns; she begins with the encrypted message, determines there are 42 characters (recall we included null letters at the end of the message, in order to obtain a rectangular matrix). The first step is to produce a 6×7 matrix:

N	R	E	N	T	A	W
L	T	E	B	I	L	R
E	O	T	D	A	D	T
B	V	A	R	H	E	E
O	R	O	M	S	T	R
X	X	X	X	O	W	X

She knows the keyword is ORANGES, which (when placed in alphabetical order) shuffles to AEGNORS. She places this at the top of the matrix as an initial "key row."

A	E	G	N	O	R	S
N	R	E	N	T	A	W
L	T	E	B	I	L	R
E	O	T	D	A	D	T
B	V	A	R	H	E	E
O	R	O	M	S	T	R
X	X	X	X	O	W	X

The columns must now be reordered so the keyword ORANGES is read across the top row. We obtain:

O	R	A	N	G	E	S
T	A	N	N	E	R	W
I	L	L	B	E	T	R
A	D	E	D	T	O	T
H	E	B	R	A	V	E
S	T	O	M	O	R	R
0	W	X	X	X	X	X

The decrypted message can now be read across the rows, disregarding the null letters appearing at the end:

"Tanner will be traded to the Braves tomorrow."

Let us now consider the case where a keyword has repeated letters.

Suppose the same message were to be sent, this time using the keyword BUNNIES. Notice (by design), it still has seven letters. The process begins as before:

B	U	N	N	I	E	S
T	A	N	N	E	R	W
I	L	L	B	E	T	R
A	D	E	D	T	O	T
H	E	B	R	A	V	E
S	T	O	M	O	R	R
0	W	X	X	X	X	X

We now determine the alphabetical order of the letters in the keyword:

B = 1
E = 2
I = 3
N = 4
N = 5
S = 6
U = 7

Observe there is now ambiguity regarding which N is being referred to as occurring 4th versus 5th. We will adopt the convention of using a hierarchy determined by the appearance of those repeated letters in the keyword.

We shuffle the columns of the matrix of the plaintext message:

B	U	N	N	I	E	S
T	A	N	N	E	R	W
I	L	L	B	E	T	R
A	D	E	D	T	O	T
H	E	B	R	A	V	E
S	T	O	M	O	R	R
0	W	X	X	X	X	X

B	E	I	N	N	S	U
T	R	E	N	N	W	A
I	T	E	L	B	R	L
A	O	T	E	D	T	D
H	V	A	B	R	E	E
S	R	O	O	M	R	T
0	X	X	X	X	X	W

The encrypted message, read down the columns and disregarding the shuffled keyword row, is:

TIAHSORTOVRXEETAOXNLEBOXNBDRMXWRTERXALDETW

Decryption of the message proceeds as in the first illustration.
Question: Using the keyword "whittier," can you decode the following message?

YHUVEFCLDUHTEHOEDLNEULAFRAEMEDMPLLATFDTPDTIBTTYOHOEARGTSS
TNTASADYALSEEAIAPBIWYMHOWOEDHGEIOTHTSRRONNEEHEGHNAIROLS
OOSRTV

Lacking the keyword, decryption here would be a challenge. There are tools we can apply, such as the "index of coincidence." The **index of coincidence** (IC or IoC) is an indicator used in cryptanalysis that makes it possible to evaluate the global distribution of letters in an encrypted message for a given alphabet.

Copious research has been applied to the distribution of letters in the English alphabet, including the work of Robert Edward Lewand (Cryptographical Mathematics) of Goucher College, Baltimore, Maryland:

Letter	Frequency
a	0.08167
b	0.01492
c	0.02782
d	0.04253

Letter	Frequency
e	0.12702
f	0.02228
g	0.02015
h	0.06094
i	0.06966
j	0.00153
k	0.00772
l	0.04025
m	0.02406
n	0.06749
o	0.07507
p	0.01929
q	0.00095
r	0.05987
s	0.06327
t	0.09056
u	0.02758
v	0.00978
w	0.0236
x	0.0015
y	0.01974
z	0.00074

Careful exploitation of these results (beyond the scope of our conversation) can assist in the decryption of ciphers such as the columnar transposition model.

Columnar transposition ciphers offer better security compared to simple transposition ciphers like the rail fence cipher. However, they can still be broken using anagramming techniques and statistical analysis.

Classical ciphers, including substitution and transposition ciphers, provide a historical perspective on the development of cryptography. While they are no longer considered secure for modern applications, they serve as the building blocks for understanding the fundamental principles of cryptography and the evolution of more advanced techniques.

13.4 Modern Cryptography

Modern cryptography has evolved from classical ciphers to address the growing need for secure communication in the digital age. This section explores the key concepts and techniques used in modern cryptography, focusing on symmetric key cryptography, including block ciphers and stream ciphers.

13.4.1 Symmetric Key Cryptography

Symmetric key cryptography, also known as secret key cryptography, uses a single key for both encryption and decryption. The sender and the recipient share the same secret key, which must be kept confidential. Symmetric key algorithms can be further classified into block ciphers and stream ciphers.

13.4.1.1 Block Ciphers (e.g., AES)

Block ciphers operate on fixed-size blocks of plaintext and ciphertext. The AES is the most widely used block cipher, which was selected by the US National Institute of Standards and Technology (NIST) in 2001. AES has a block size of 128 bits and supports key sizes of 128, 192, and 256 bits.

AES uses a substitution-permutation network (SPN) structure, which consists of a series of rounds that include the following operations:

- SubBytes: a nonlinear substitution step that operates on each byte of the state matrix independently.
- ShiftRows: a transposition step that cyclically shifts the rows of the state matrix.
- MixColumns: a linear mixing operation that combines the bytes in each column of the state matrix.
- AddRoundKey: a bitwise XOR operation that combines the state matrix with the round key.

Example 13.3 Encrypting the plaintext "Hello, World!" using AES-128 with the key "0123456789 abcdef" results in the ciphertext "0x7c 0x87 0x2d 0x6d 0x5e 0xf1 0x8a 0xb1 0x25 0xf8 0xc3 0x5d 0x68 0x89 0xd1 0x53".

Example 13.4 Consider a simplified version of AES with a block size of 16 bits and a key size of 16 bits. Let the plaintext be "0110 1001 1100 0011" and the key be "1010 0101 0011 1100". After applying the AES operations (SubBytes, ShiftRows, MixColumns, and AddRoundKey) for a single round, the resulting ciphertext would be "1101 0011 0110 1001".

Other notable block ciphers include:

- DES (Data Encryption Standard): DES was once widely used but has been largely deprecated due to its small key size of 56 bits, which makes it vulnerable to brute-force attacks.
- Blowfish: Blowfish is a block cipher designed by Bruce Schneier in 1993. It has a variable key size ranging from 32 to 448 bits and is known for its fast encryption speed.
- IDEA (International Data Encryption Algorithm): IDEA is a block cipher that operates on 64-bit blocks and uses a 128-bit key. It was designed by Xuejia Lai and James Massey in 1991.

13.4.1.2 Stream Ciphers (e.g., RC4)

A stream cipher is an encryption technique that works byte by byte to transform plaintext into code unreadable to anyone without the proper key. Stream ciphers are linear, so the same key serves to both encrypt and decrypt messages.

While breaking such an encryption can be challenging, hackers have managed to do it; for that reason, stream ciphers aren't safe for widespread use in the transmission of highly sensitive information. Nevertheless, many people lean on technology to pass information (with or without recognizing the risks in doing so) through the internet.

A historically significant use of stream ciphers is the German encryption device known as the Enigma Machine, used by officials to send directions to troops in the field.

The virtues of stream ciphers are manifold:

- **Speed.** This form of encryption is typically faster than others, including block ciphers.
- **Low complexity.** Stream ciphers are easy to incorporate into modern programs, without the need for complex hardware.
- **Serial nature.** Stream ciphers allow them to send information when it is ready rather than waiting for everything to be done.
- **Ease of use.** Stream ciphers are symmetrical encryption tools, so companies are not required to employ both public and private keys.

Stream ciphers encrypt plaintext bits individually by combining them with a pseudorandom keystream that exhibits enhanced chaos properties to frustrate efforts at decryption.

The keystream is generated using a secret key and an initialization vector (IV). An IV is an arbitrary number that can be used with a secret key for data encryption to foil cyber-attacks. Also called a **nonce** (number used once), an IV is employed only a single time in any session, to prevent unauthorized decryption of the message by a suspicious or malicious actor. One of the most widely used stream ciphers is Rivest Cipher 4 (RC4), which was designed by Ron Rivest in 1987.

RC4 generates a keystream by using a secret key to initialize a 256-byte state array (a four-by-four matrix of variable values), which is then permuted and used to generate pseudorandom bytes. The plaintext is encrypted by performing a bitwise XOR operation with the keystream.

Example 13.5 Encrypting the plaintext "Hello, World!" using RC4 with the key "secret" results in the ciphertext "0xeb 0x0a 0x0c 0xbc 0xf9 0xd7 0x42 0x5a 0xab 0xf3 0xd2 0x40".

Example 13.6 Consider a simplified RC4-like stream cipher with a 4-bit key and a 4-bit plaintext. Let the key be "1011" and the plaintext be "0110". The keystream is generated by permuting the initial state array using the key and then XORing the plaintext with the keystream. The resulting ciphertext would be "1101".

It is worth mentioning that RC4 suffers from biases in its key scheduling algorithm. This can lead to statistical biases in the generated keystream, an obscure but relevant weakness a hacker can exploit to deduce information about the key sufficient to recover part or all of the plaintext.

Other notable stream ciphers include:

- ChaCha20: ChaCha20 is a modern stream cipher designed by Daniel J. Bernstein. It is used in various protocols, including Transport Layer Security (TLS) and Quick User Datagram Protocol (UDP) Internet Connections (QUIC).
- A5/1: A5/1 is a stream cipher that was used for over-the-air communication privacy in Global System for Mobile Communications (GSM) cellular networks. However, it has been shown to have vulnerabilities and is no longer considered secure.
- Salsa20: Salsa20 is a stream cipher designed by Daniel J. Bernstein in 2005. It has a 256-bit key and a 64-bit nonce, and it is known for its simplicity and fast software implementation.

Symmetric key cryptography, particularly block ciphers and stream ciphers, forms the foundation of modern cryptography. AES and RC4 are widely used examples of block ciphers and stream ciphers, respectively. While they provide efficient and secure encryption, it is important to use them correctly and with appropriate key management to ensure the overall security of the cryptographic system.

13.4.2 Public Key Cryptography

Public key cryptography, also known as asymmetric cryptography, uses a pair of keys: a public key for encryption and a private key for decryption. It is the dual nature of the keys that prompts the name "asymmetric cryptography." The public key can be freely distributed, while the private key is kept secret by the owner. Public key cryptography enables secure communication without the need for a pre-shared secret key.

13.4.2.1 RSA Cryptosystem

The RSA cryptosystem is one of the most widely used public key algorithms. It relies on the mathematical difficulty of factoring large composite numbers. The RSA algorithm consists of the following steps:

1) Key Generation:
   - Choose two large prime numbers, p and q.
   - Compute $n = p \times q$ and $\varphi(n) = (p-1) \times (q-1)$ [the latter expression is called the "totient"].
   - Choose an integer e such that $1 < e < \varphi(n)$ and gcd(e, $\varphi(n)$) = 1.
   - Compute d such that $d \times e \equiv 1 \pmod{\varphi(n)}$.
   - The public key is (n, e), and the private key is (n, d).
2) Encryption:
   - Represent each character of the plaintext message as an integer m, where $0 \leq m < n$.
   - Compute the ciphertext $c = m\text{^}e \bmod n$.
3) Decryption:
   - Compute the plaintext $m = c\text{^}d \bmod n$.

For instance, let $p = 17$, $q = 11$, and $e = 7$. One can debate whether p and q are "large," but let us set this aside. We find $n = (p)(q) = 187$, and $\varphi(187) = (16)(10) = 160$.

Next, choose an integer e such that $1 < e < 160$, with gcd(e,160) = 1. Some explanation is needed here; "gcd" refers to "greatest common divisor," and the criterion gcd(e,160) = 1 is described as e and 160 being "relatively prime." We can pick any number for e that satisfies these conditions. The thing to do is find the prime factorization of $\varphi(187) + 1$, which here is (7)(23), and choose as d any of these prime factors of $\varphi(187) + 1$.

For the sake of example, we select $d = 23$. This tells us $e = 7$. In practice, the values of e that are typically chosen are among 3, 17, and 65,537. The selection of "e" in the RSA algorithm depends on the values of "p" and "q" and the requirement that "e" must be coprime to the totient of n. The value of "e" can, in fact, be any number that meets this requirement.

> *NOTE: The choice of 65 337 may seem unusual, but it is, in fact, $2^{16} + 1$, which has prime factorization (3)(29)(751).*

The public key is (n, e) = (187, 7), and the private key is (n, d) = (187, 23).

A reasonable question is: how does this relate to alphabetical characters? This would seem to be of paramount importance, else how would we make sense of transmission? Any relationship between letters and numbers would suffice, but a common selection is to use the American Standard Code for Information Interchange (ASCII) code for letters:

A	B	C	D	E	F	G	H	I	J	K	L	M
65	66	67	68	69	70	71	72	73	74	75	76	77
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
78	79	80	81	82	83	84	85	86	87	88	89	90

In this setup, a letter such as "H" would correspond to the two-digit representation "72." As you can see, we have a relationship between all the letters of the alphabet and a numerical value. This, of course, is limited to the alphabetical relations for ASCII (in decimal form) and is incomplete regarding all other characters. A slightly enlarged table would be:

Dec	Char	Dec	Char	Dec	Char	Dec	Char
0	Ctrl-@ NUL	32	Space	64	@	96	`
1	Ctrl-A SOH	33	!	65	A	97	a
2	Ctrl-B STX	34	"	66	B	98	b
3	Ctrl-C ETX	35	#	67	C	99	c
4	Ctrl-D EOT	36	$	68	D	100	d
5	Ctrl-E ENQ	37	%	69	E	101	e
6	Ctrl-F ACK	38	&	70	F	102	f
7	Ctrl-G BEL	39	'	71	G	103	g
8	Ctrl-H BS	40	(	72	H	104	h
9	Ctrl-I HT	41	)	73	I	105	i
10	Ctrl-J LF	42	*	74	J	106	j
11	Ctrl-K VT	43	+	75	K	107	k
12	Ctrl-L FF	44	,	76	L	108	l
13	Ctrl-M CR	45	-	77	M	109	m
14	Ctrl-N SO	46	.	78	N	110	n
15	Ctrl-O SI	47	/	79	O	111	o
16	Ctrl-P DLE	48	0	80	P	112	p
17	Ctrl-Q DCI	49	1	81	Q	113	q
18	Ctrl-R DC2	50	2	82	R	114	r
19	Ctrl-S DC3	51	3	83	S	115	s
20	Ctrl-T DC4	52	4	84	T	116	t
21	Ctrl-U NAK	53	5	85	U	117	u
22	Ctrl-V SYN	54	6	86	V	118	v
23	Ctrl-W ETB	55	7	87	W	119	w
24	Ctrl-X CAN	56	8	88	X	120	x
25	Ctrl-Y EM	57	9	89	Y	121	y
26	Ctrl-Z SUB	58	:	90	Z	122	z
27	Ctrl-[ESC	59	;	91	[	123	{
28	Ctrl- FS	60	<	92	₩	124	¦
29	Ctrl-] GS	61	=	93	]	125	}
30	Ctrl-^ RS	62	>	94	^	126	~
31	Ctrl_ US	63	?	95	_	127	DEL

This is not complete, either, but it gives one the idea of what digits we would use as a tool for the encryption of alphanumeric characters.

RSA security relies on the difficulty of factoring large numbers. It is important to choose sufficiently large prime numbers (e.g., 2048 bits or more) to ensure the security of the cryptosystem.

13.4.2.2 Diffie-Hellman Key Exchange

The Diffie-Hellman key exchange, proposed by Whitfield Diffie and Martin Hellman in 1976, is a method for securely establishing a shared secret key over an insecure channel. That is, the process allows cryptographic keys to be exchanged safely in "plain sight."

Though we are only viewing this technique from "on high," not delving into its derivation or complexities, the briefly described premise of the exchange is staggering in its importance. Diffie and Hellman asserted their result placed us "on the brink of a revolution in cryptography," and time has shown this was no exaggerated claim.

Let us restate the issue this technique brilliantly addresses: can two parties in the process of transmitting information establish a secure channel of communication and verify one another's identity? As you can deduce (because we are presenting the method), the answer to this question is "yes."

Doing so relies upon using a "one-way function." A one-way function is easy to compute in one direction but hard to reverse. This means you can quickly get the output from an input, but it is a challenge to discern the input from the output. Using such a function, the parties involved can establish a secret number each of them knows, but that a bad actor attempting to intercept the information cannot deduce. Once established, the users can switch to a relatively simple symmetric encryption to quickly transmit their information to one another.

The key exchange relies on the difficulty of computing discrete logarithms in a "finite field," also known as a "Galois field." Mathematically, a finite (or Galois) field is a finite set of objects on which the operations of multiplication, addition, subtraction, and division are defined and satisfy certain basic rules (full exploration of this concept lies in the province of Abstract Algebra and Galois Theory; here we are giving only a thumbnail sketch).

Imagine the following scenario: Alfred and Nadine wish to speak to one another privately, but realize Stephan has access to their communications. Assuming Alfred and Nadine have not agreed, in advance, upon some secret transmission key unknown to Stephan, how can they talk to one another securely?

If Alfred and Nadine could construct a one-way function through which they could exchange a key (so Stephan couldn't deduce it, even though he has access to their messages), such a private conversation could be conducted. This necessitated an exchange of some basic information between Alfred and Nadine; they were willing to let any observer (including Stephan) see; this would serve as a "public key."

Behind the scenes, the one-way function would allow for the creation of a secret key about which no one else would be aware. Because of the existence of both a public and private key, this technique is called an "asymmetric encryption technique."

The Diffie-Hellman key exchange consists of the following steps:

1) Those interested in establishing secure communications begin with a function of the form $g^a (mod\ P)$, where the values of g and P are agreed upon, and P is prime. In theory, P could be any prime, but as a matter of practical security, some large-valued prime [having more than 617 digits]. The larger the number of digits, the more resistant the exchange is to external attack.
2) Those interested in secure communications (Persons A and B) each select a secret integer and exchange those values with one another. Person A computes $A = g^{integer} (mod\ P)$, Person B computes $B = g^{integer} (mod\ P)$, and share their respective values (A and B) with one another.
3) Person A computes the shared secret key $K = B^a \pmod P$, while Person B computes the shared secret key $K = A^b \pmod P$. This shared key K can now be used in symmetric key encryption.

The remarkable thing about this process is that **both of the secret keys are the same**; just why this is the case is unclear without further exploration, but a vital component in the situation is that the modulus P is prime.

Without proof, let us demonstrate the values of K will be the same for each person, using relatively small values for the sake of example.

Suppose Alfred and Nadine agree to use the base value $g = 11$ and the prime to be $P = 13$. The public function, known to everyone (including Stephan), is now $11^a (mod\ 13)$. Alfred and Nadine select values of $a = 5$ and 8, respectively, and let one another know these values.

Alfred's function (known only to him) is $11^5 (mod\ 13) \equiv 7 (mod\ 13)$, Nadine's is $11^8 (mod\ 13) \equiv 9$ $(mod\ 13)$, and Stephan is only aware of $11^a (mod\ 13)$.

Alfred calculates his version of the secret key $K = 9^5 (mod\ 13) = 59049 (mod\ 13) = 3 (mod\ 13)$; Nadine calculates her version of $K = 7^8 (mod\ 13) = 5764801 (mod\ 13) = 3 (mod\ 13)$.

This is not a proof, but observe the enormity of the numbers involved. Despite the incredible size (even with small values of g and P) of the values, the value of K thus generated is the same for both Alfred and Nadine.

Using a brute-force technique of trial and error, Stephan could (given time) crack this cipher without significant difficulty. However, keep in mind the intention was to use a much larger value of P, having at least 617 digits. Attempting to perform the computations above in such a case would be a daunting task, calling for the solution of what we call the "discrete logarithm problem."

Using their private key, Alfred and Nadine are able to create and use a symmetric encryption technique, while Stephan would be confronted with an enormous problem to solve, requiring a massive expenditure of computer time and capital.

Parenthetically, we remark that it is common to use 2048 encryption in the current working environment, which would require Stephan to sort through 2^{2048} possible key values. This is a bewilderingly large number, almost beyond comprehension. As a frame of reference 2^{100}, which is much smaller than the number of possibilities Stephan faces, is 1,267,650,600,228,230,000,000,000,000,000—more than 1 nonillion possible numbers.

Example 13.7 Let $p = 23$ and $g = 5$. Alice chooses $a = 6$ and computes A = 5^6 mod 23 = 8. Bob chooses $b = 15$ and computes B = 5^15 mod 23 = 19. Alice computes K = 19^6 mod 23 = 2, and Bob computes K = 8^15 mod 23 = 2. The shared secret key is 2.

Example 13.8 Consider a larger Diffie-Hellman instance with $p = 2147483647$ (a prime number) and $g = 16\,807$. Alice chooses $a = 1234567890$ and computes A = 16807^1234567890 mod 2147483647 = 1073741823. Bob chooses $b = 987654321$ and computes B = 16807^987654321 mod 2147483647 = 1342177280. Alice computes K = 1342177280^1234567890 mod 2147483647 = 511875698, and Bob computes K = 1073741823^987654321 mod 2147483647 = 511875698. The shared secret key is 511875698.

The security of the Diffie-Hellman key exchange relies on the discrete logarithm problem. Determining the secret integers a and b is computationally infeasible, given the public values A, B, p, and g.

Public key cryptography, particularly the RSA cryptosystem and the Diffie-Hellman key exchange, revolutionized secure communication by enabling secure key exchange and encryption without the need for a pre-shared secret key. These techniques form the basis for many modern cryptographic protocols, such as Secure Sockets Layer (SSL)/TLS and Secure Shell (SSH), which are widely used for online secure communication.

13.4.3 Hash Functions and Message Authentication Codes

Hash functions (processes that take plaintext data of any size and convert it into a unique ciphertext of a specific length) and message authentication codes (MACs) (used to authenticate the origin and nature of a message) are cryptographic primitives used to ensure data integrity and authenticity.

13.4.3.1 SHA-256 and SHA-3

A hash function is a one-way function that maps an arbitrary-length input to a fixed-length output called a hash or digest. Cryptographic hash functions should be collision-resistant, meaning that it is computationally infeasible to find two distinct inputs that produce the same hash value.

Secure Hash Algorithm 256-bit (SHA-256) is a widely used hash function that produces a 256-bit digest. It is part of the SHA-2 family of hash functions designed by the US National Security Agency (NSA). SHA-256 uses a Merkle-Damgård construction and is based on a series of bitwise operations and modular additions.

Example 13.9 The SHA-256 hash of the message "Hello, World!" is "0x7f 0x83 0xb1 0x65 0x7f 0xf1 0x6a 0xde 0xb8 0xe1 0xee 0x99 0x2b 0x3e 0x9f 0x5d 0x2b 0x5b 0x23 0x3c 0x87 0x62 0xb6 0x2e 0x65 0xff 0x59 0xa3 0x0d 0x75 0x98".

Example 13.10 Consider two similar messages: "The quick brown fox jumps over the lazy dog" and "The quick brown fox jumps over the lazy cog". Despite only a single character difference, their SHA-256 hashes are significantly different:

- "The quick brown fox jumps over the lazy dog" → "0xd7 0xa8 0xfb 0xb3 0x07 0xd7 0x80 0x94 0x69 0xca 0x9a 0xbc 0xb0 0x08 0x2e 0x4f 0x8d 0x56 0x51 0xe4 0x6d 0x3c 0xdb 0x76 0x2d 0x02 0xd0 0xbf 0x37 0xc9 0xe5"
- "The quick brown fox jumps over the lazy cog" → "0xe4 0xc0 0x9c 0x6d 0x6f 0x75 0xc5 0xec 0x07 0x8a 0x0d 0xb7 0x1b 0xd4 0x3a 0x8b 0x87 0xf3 0xd0 0x4b 0x20 0x7b 0x11 0xd8 0x9e 0x85 0xca 0xf7 0xd2 0xa2 0xd3 0x7c"

SHA-3 is a more recent hash function standard selected by NIST in 2012. It is based on the Keccak algorithm, which uses a "sponge construction." The sponge construction is a simple iterated construction building a variable-length input, variable-length output function based on a fixed-length permutation. Full exploration of this notion, as with many advanced topics we have mentioned, lies beyond the scope of this work. SHA-3 offers several variants with different output sizes, including SHA3-224, SHA3-256, SHA3-384, and SHA3-512.

Example 13.11 The SHA3-256 hash of the message "Hello, World!" is "0x64 0x4b 0xcc 0x7e 0x56 0x43 0x73 0x04 0x09 0x99 0xaa 0xc8 0x9e 0x76 0x22 0xf3 0xca 0x71 0xfb 0xa1 0xd9 0x72 0xfd 0x94 0xa3 0x1c 0x3b 0xfb 0xf2 0x4e 0x39 0x38".

13.4.3.2 HMAC

Hash-based Message Authentication Code (HMAC) is a type of MAC that uses a cryptographic hash function combined with a secret key. HMACs provide data integrity and authentication, ensuring that the message has not been altered and that it originates from the claimed sender.

A MAC is like a cryptographic hash, except that it is based on a secret key. When secret key information is included with the data processed by a cryptographic hash function, the resulting hash is referred to as an HMAC.

The HMAC algorithm works as follows:

1) Append zeros to the secret key to make it the same size as the hash function's block size.
2) XOR the padded key with the outer padding (0x5c5c…5c).
3) Append the message to the result of step 2.
4) Apply the hash function to the result of step 3.
5) XOR the padded key with the inner padding (0x3636…36).
6) Append the hash result from step 4 to the result of step 5.
7) Apply the hash function to the result of step 6 to produce the final HMAC value.

Example 13.12 Consider the message "Hello, World!" and the secret key "secret". The HMAC-SHA256 value is computed as follows:

- Padded key (256 bits): "0x73 0x65 0x63 0x72 0x65 0x74 0x00 0x00 … 0x00" (padded with zeros)
- Outer padding: "0x5c 0x5c 0x5c 0x5c 0x5c 0x5c 0x5c 0x5c … 0x5c"
- Inner padding: "0x36 0x36 0x36 0x36 0x36 0x36 0x36 0x36 … 0x36"
- Intermediate hash: SHA256((padded_key XOR outer_padding) || message)
- Final HMAC: SHA256((padded_key XOR inner_padding) || intermediate_hash)

The resulting HMAC-SHA256 value is "0x49 0x2a 0x72 0xd9 0x3e 0x31 0x19 0xeb 0xb1 0xa1 0x5a 0x14 0x8b 0x6f 0x62 0x35 0xed 0x1b 0x3d 0x66 0x14 0x2e 0x9d 0x0a 0xf4 0x88 0x52 0x98 0xbf 0x4b 0x55 0xfa".

Example 13.13 Consider the message "The quick brown fox jumps over the lazy dog" and the secret key "key". The HMAC-SHA3-256 value is computed similarly to Example 13.12, but using the SHA3-256 hash function. The resulting HMAC-SHA3-256 value is "0x6a 0x4b 0x27 0xb6 0x98 0xc2 0xa7 0x69 0x91 0x09 0x71 0x31 0x9e 0x48 0x1b 0xe6 0xf6 0x28 0x3c 0xd5 0xc0 0xe6 0x5c 0x7c 0x50 0x9a 0x1f 0xf3 0x9d 0x9e 0x17 0x33".

13.5 Applications of Cryptography

13.5.1 Secure Communication

Communication is considered "secure" when information can be transmitted confidentially between the participants over potentially insecure channels. As we have seen, there exist numerous techniques (such as symmetric and public key encryption) to protect the confidentiality of messages.

Example 13.14 HTTPS (Hypertext Transfer Protocol Secure) is a widely used protocol for secure communication on the internet. Not to be mistaken as the plural of HTTP, which is the foundation upon which the "World Wide Web" is built. It is used to load webpages using hypertext links. HTTP is an application layer protocol designed to transfer information between networked devices and runs on top of other layers of the network protocol stack.

When a user accesses a website using HTTPS, the communication between the user's browser and the web server is encrypted using SSL/TLS (Secure Sockets Layer). The encryption prevents eavesdroppers from intercepting and reading the transmitted data, such as login credentials or sensitive information, and is an instantiation of an asymmetric security protocol.

This type of security system uses two different keys to encrypt communications between two parties:

1) The private key: this key is controlled by the owner of a website, and it is kept, as the reader may have speculated, private. This key lives on a web server and is used to decrypt information encrypted by the public key.
2) The public key: this key is available to everyone who wants to interact with the server in a way that is secure. Information that is encrypted by the public key can only be decrypted by the private key.

HTTPS prevents websites from having their information broadcast in a way that's easily viewed by anyone snooping on the network. When information is sent over regular HTTP, the information is broken into packets of data that can be easily "sniffed" using free software. This makes communication over an insecure medium (such as public Wi-Fi available in a restaurant) highly vulnerable to interception. In fact, all communications that occur over HTTP occur in plain text, making them highly accessible to anyone with the correct tools.

Example 13.15 Secure messaging applications, such as Signal and WhatsApp, use end-to-end encryption (E2EE) to protect the privacy of user conversations. The messages are encrypted on the sender's device and can only be decrypted by the intended recipient. This ensures that even if an attacker intercepts the messages, they cannot read the contents without the encryption key.

This is important because (without it), while your information may be encrypted when being sent to the server, it is possible the server itself might be able to read it. For example, some service providers might do this to generate ads that are more specific to a user.

When you are browsing a social media site, you may notice the appearance of advertisements that seem almost eerily personalized toward your interests. This is not because the computer system has some form of mind-reading capability, but rather that the server to which you are connected is able to analyze your content and tailor advertisements toward you.

13.5.2 Data Integrity and Authentication

Cryptography is also used to ensure the integrity and authenticity of data. **Integrity** means the data has not been altered or tampered with on its way to you, while authenticity ensures that the data originates from a trusted source. Hash functions and MACs are commonly used for this purpose.

13.5.3 Digital Signatures

Digital signatures are a cryptographic technique providing authentication, non-repudiation, and integrity for digital documents or messages. A digital signature is a mathematical scheme that uses public key cryptography to verify a message's authenticity and the signer's identity.

Cryptocurrencies, like Bitcoin and Ethereum, heavily rely on digital signatures for securing transactions. Each transaction is digitally signed by the owner of the funds using their private key. The network verifies the signature using the corresponding public key to ensure the authenticity and integrity of the transaction. This prevents unauthorized spending and double-spending of funds.

Using digital signatures in conjunction with Public Key Infrastructure (PKI) or PGP strengthens them and reduces the possible security issues connected to transmitting public keys. This is achieved by validating the key belongs to the legitimate sender and verifying the identity of that

sender. When a third party serves as an intermediary in the process of communication, digital signatures can be used to identify and verify individuals and ensure the integrity of the message.

The security of a digital signature is almost entirely dependent on how well the private key is protected. Without PGP or PKI, proving someone's identity or revoking a compromised key is impossible; this could allow malicious actors to impersonate someone without any method of confirmation.

13.6 Cryptanalysis and Cryptographic Attacks

Cryptanalysis is the study of techniques used to break cryptographic systems and gain unauthorized access to encrypted information. There are various types of cryptanalytic attacks, including brute force attacks, frequency analysis, and known plaintext and chosen plaintext attacks. We now give a brief overview of these techniques.

13.6.1 Brute Force Attacks

A brute force attack is a cryptanalytic technique that involves exhaustively trying all possible keys or combinations until the correct one is found. This type of attack is based on the attacker's computational power and the size of the key space.

Imagine two individuals attempting to break into a home: the first takes a sledgehammer and smashes the door lock, enters the building, and commits a burglary; the second uses an elaborate lockpicking technique, eases open the door, enters the building, and commits the burglary. Though both methods are illegal practices, we can distinguish between the two of them in terms of sophistication. The first, use of the sledgehammer, we can recognize as an application of brute force. The second is an application of greater skill and technique.

Example 13.16 Consider a simple substitution cipher with a key space of 26! (factorial of 26), which is approximately 4×10^{26}. If an attacker can try 1 million keys per second, it would take about 10^{20} seconds, or 3×10^{12} years, to exhaust the entire key space. This is computationally infeasible, making the brute force attack impractical for this cipher.

Example 13.17 In contrast, consider a 4-digit PIN (Personal Identification Number) used for ATM access. The key space is only 10,000 possibilities (0000 to 9999). An attacker with a stolen ATM card can try all possible PINs in a relatively short time, making the system vulnerable to brute force attacks.

Moreover, individuals are notoriously inept at attempting to create random passwords on their own, with (in this specific instance) many people relying upon highly predictable Personal Identification Number (PIN) choices; by far, the most frequently used four-digit PIN in use is "1234."

Research by DataGenics, a company whose specialty is cybersecurity, has shown that (statistically) the top 20 preferred PINs comprise more than 25% of all those in use, and that the top 426 preferred PINs (out of the possible 10 000 alternatives) represent more than 50% of all PINS being used currently.

To defend against brute force attacks, it is essential to use sufficiently large key spaces and implement measures such as throttling or locking accounts after a certain number of failed attempts.

13.6.2 Frequency Analysis

Frequency analysis is a technique used to break classical ciphers, such as substitution ciphers, by analyzing the frequency of letters or symbols in the ciphertext. This method exploits the statistical properties of the language used in the plaintext.

Example 13.18 We have mentioned that, in the English language, the letter 'e' is the most frequently occurring letter, followed by "t", "a", "o", and "i". In a simple substitution cipher, if the most frequent letter in the ciphertext is "X", it is likely that "X" represents the plaintext letter "e".

Example 13.19 Consider the following ciphertext: "QAWVR ZAQEB VGURA EBVFR QVALA TVARF". By analyzing the frequency of the letters, we can deduce that "R" (occurring 6 times) likely represents "e", "A" (occurring 4 times) represents "t" or "a", and so on. By substituting the letters based on frequency analysis, we can decrypt the ciphertext to "NEVER GONNA GIVE YOU UP NEVER GONNA LET YOU DOWN".

To defend against frequency analysis, techniques like polyalphabetic substitution (e.g., Vigenère cipher) or homophonic substitution can be used to "flatten" the frequency distribution of the ciphertext.

13.6.3 Known Plaintext and Chosen Plaintext Attacks

Known plaintext and chosen plaintext attacks are cryptanalytic techniques that exploit the knowledge of some plaintext-ciphertext pairs to deduce the encryption key or algorithm.

In a known plaintext attack, the attacker has access to a set of plaintext-ciphertext pairs encrypted with the same key. By analyzing these pairs, the attacker attempts to determine the key or the encryption algorithm.

Example 13.20 In a simple XOR cipher, if the attacker knows the plaintext "HELLO" and its corresponding ciphertext "0x2d 0x11 0x1a 0x0e 0x0f", they can XOR the plaintext and ciphertext bytes to recover the key "0x48 0x45 0x4c 0x4c 0x4f".

In a chosen plaintext attack, the attacker can choose arbitrary plaintext messages and obtain their corresponding ciphertexts. This provides the attacker with more control and flexibility in analyzing the cryptographic system.

Example 13.21 In a block cipher, an attacker can exploit a chosen plaintext attack by encrypting a large number of plaintext blocks that differ in specific bits. By analyzing the corresponding ciphertext blocks, the attacker can deduce the key bits or the internal structure of the block cipher.

To defend against known plaintext and chosen plaintext attacks, modern cryptographic algorithms, such as AES and RSA, are designed to be secure against these types of attacks. Techniques like key randomization, message padding, and encryption modes (e.g., Cipher Block

Chaining [CBC], Counter Mode [CTR]) are used to mitigate the impact of known or chosen plaintext attacks.

Cryptanalysis and cryptographic attacks are essential considerations when designing and evaluating cryptographic systems. Understanding techniques like brute force attacks, frequency analysis, and known/chosen plaintext attacks helps in developing more secure algorithms and protocols. By employing strong key sizes, randomization techniques, and well-studied algorithms, cryptographic systems can be made more resilient against these types of attacks.

13.7 Conclusion and Future Directions

As we have seen throughout this chapter, cryptography plays a vital role in ensuring the security and privacy of information in the digital age. From classical ciphers to modern symmetric and asymmetric algorithms, cryptography has evolved to meet the growing challenges of secure communication and data protection. In this final section, we will explore two emerging areas that represent the future directions of cryptography: quantum cryptography and post-quantum cryptography.

13.7.1 Quantum Cryptography

Quantum cryptography is a fascinating area that combines principles from quantum mechanics and cryptography to provide unconditionally secure communication. The most well-known application of quantum cryptography is Quantum Key Distribution (QKD).

QKD enables two parties, usually referred to as Alice and Bob, to establish a shared secret key over a quantum channel. The security of QKD is based on the fundamental principles of quantum mechanics, such as the Heisenberg uncertainty principle and the no-cloning theorem.

Example 13.22 The BB84 protocol is one of the most well-known QKD protocols. In this protocol, Alice encodes random bits into the polarization states of photons (e.g., horizontal, vertical, diagonal, or anti-diagonal) and sends them to Bob. Bob randomly chooses to measure each photon's polarization in either the rectilinear (horizontal or vertical) or diagonal (diagonal or anti-diagonal) basis. After the transmission, Alice and Bob compare their bases and discard the bits where their bases differ. The remaining bits form the raw key, which is then processed through error correction and privacy amplification to obtain the final secret key.

Example 13.23 The E91 protocol, proposed by Artur Ekert, uses entangled photon pairs to establish a shared secret key. Alice and Bob each receive one photon from an entangled pair and perform measurements in random bases. By comparing their measurement results, they can detect any eavesdropping attempts and establish a secure key.

Quantum cryptography offers the unique advantage of detecting eavesdropping attempts, as any measurement performed by an eavesdropper on the quantum channel will introduce detectable errors. However, practical implementation of quantum cryptography still faces challenges, such as the limited distance of quantum channels and the need for specialized hardware.

13.7.2 Post-quantum Cryptography

Post-quantum cryptography, also known as quantum-resistant cryptography, refers to the development of cryptographic algorithms that are secure against attacks by quantum computers.

With the rapid advancement of quantum computing, there is a growing concern that quantum computers may be able to break some of the widely used public-key cryptographic algorithms, such as RSA and ECC, in the near future. This has led to the search for post-quantum cryptographic algorithms that can withstand attacks by both classical and quantum computers.

Example 13.24 Lattice-based cryptography is a promising candidate for post-quantum cryptography. Lattice-based schemes, such as the Learning with Errors (LWE) problem and the Ring Learning with Errors (RLWE) problem, are believed to be secure against quantum attacks. The security of these schemes is based on the hardness of solving certain mathematical problems in high-dimensional lattices.

Example 13.25 Hash-based digital signature schemes, such as the Merkle signature scheme and the Leighton-Micali signature scheme, are another class of post-quantum cryptographic algorithms. These schemes rely on the security of hash functions and are designed to be resistant to quantum attacks.

The NIST is currently conducting a standardization process for post-quantum cryptographic algorithms. The goal is to select and standardize quantum-resistant public-key encryption, digital signature, and key establishment algorithms.

As quantum computers become more powerful, it is crucial to transition to post-quantum cryptographic algorithms to ensure the long-term security of sensitive information and communications.

Cryptography is an ever-evolving field that must adapt to new challenges and technological advancements. Quantum cryptography and post-quantum cryptography represent two exciting areas that aim to address the unique security challenges posed by quantum computing. As we move forward, it is essential for researchers, practitioners, and policymakers to collaborate and work toward developing and implementing secure and practical cryptographic solutions that can withstand the test of time.

13.8 A Closing Remark on a Nearly Unbreakable Code, Resolved After 50 Years

During the late 1960s, at least five individuals were murdered in the United States by a serial killer who used the alias “Zodiac.” The final killing purportedly was committed in October of 1969, and the murderer (who was never identified) claimed in a message to the police to have killed 37 people.

The Zodiac Killer taunted police in San Francisco by sending messages to them using elaborate ciphers, the first of which (known as the Z408 cipher, because the message contained 408 characters) was published by a group of local newspapers: the Vallejo Times-Herald, the San Francisco Examiner, and the San Francisco Chronicle, each of which had been sent one-third of the message. Eight days after publication, the cipher was broken at a breakfast table by Donald and Bettye Harden—a couple living in nearby Salinas, California.

Shown is an image of the Z408 message:



Source: Image from Federal Bureau of Investigation / Public Domain.

Zodiac Killer Z408 cipher
The key to this cipher is:

A6SJ▲ BV C∃ D⊽✦ EZ9W+⊙NE FJQ
GR HM⊷ IΔPUK K/ L◪B■ MΘ
NO∧DΦ OXIT9 P⊼ RL⅄\ SF⊡K△
THI●L UY V⊃ WA XT Y□

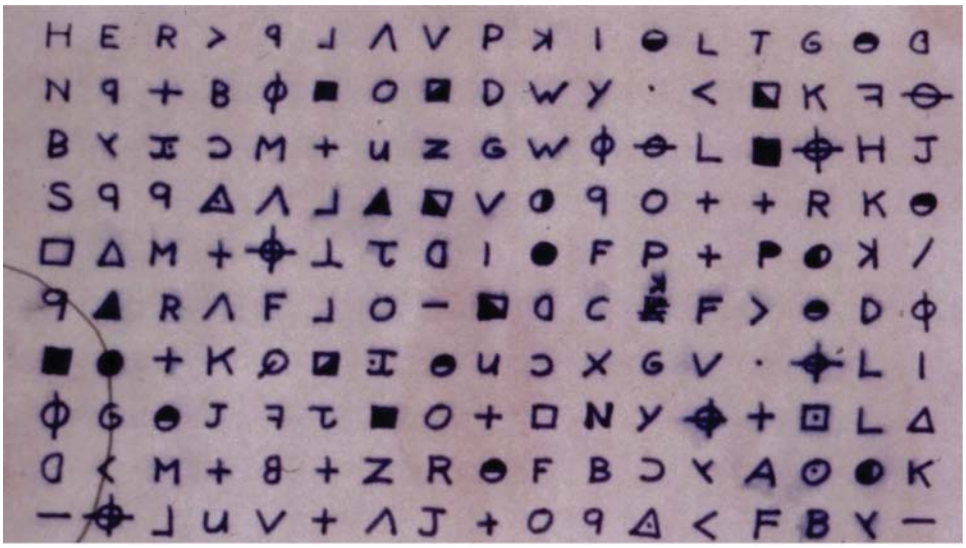
Observe multiple symbols are used for many of the letters of the alphabet; there are, for instance, four symbols equivalent to the letter "A." The plaintext of the first third of this message is:



Observe the ominous manner in which the message begins: "I like killing people."

Later ciphers sent by the Zodiac Killer have baffled cryptanalysts for decades, one of which (again named for its number of characters) was called the Z340 cipher. This cipher (made more complicated by writing the words of the encryption diagonally with occasional columnar shifts and some misspelling of words) was defeated in 2021 through the collaboration of David Oranchak (a web designer in Virginia), Sam Blake (a mathematician in Australia), and Jarl Van Eykcke (a warehouse operator in Belgium).

As released by the team and the Federal Bureau of Investigation, the solution to the cipher (shown below an image of the original message) reads:



Source: Image from Federal Bureau of Investigation / Public Domain.

I HOPE YOU ARE HAVING LOTS OF FUN IN TRYING TO CATCH ME THAT WASNT ME ON THE TV SHOW WHICH BRINGS UP A POINT ABOUT ME I AM NOT AFRAID OF THE GAS CHAMBER BECAUSE IT WILL SEND ME TO PARADICE ALL THE SOONER BECAUSE I NOW HAVE ENOUGH SLAVES TO WORK FOR ME WHERE EVERYONE ELSE HAS NOTHING WHEN THEY REACH PARADICE SO THEY ARE AFRAID OF DEATH I AM NOT AFRAID BECAUSE I KNOW THAT MY NEW LIFE IS LIFE WILL BE AN EASY ONE IN PARADICE DEATH.

The reference to a television show concerns a call to "The Jim Dunbar Show" in San Francisco, in which a caller claimed to be the Zodiac Killer.

We close with this application of cryptanalysis as a means to highlight that advancements in computing technology continue to make problems once thought to be intractable solvable, and to point out the high relevance of the notions presented in this work.

Index

a

Absorption laws 29
Addition rule 349
Adjacency list 390, 409–411
Adjacency matrix 387–390, 397, 408–411
Adjacent vertices 382, 385–386
Advanced Encryption Standard (AES) 88, 446
Advanced mathematical concepts in security 88, 446
AES encryption 88
Algorithm efficiency 404, 443
Algorithms 18, 161, 195, 204, 259–291, 295, 378, 380, 398, 399, 410, 419, 420, 422, 426, 432–444, 462, 470, 472
Alice and Bob examples 384, 395, 471
Angle measurement 207, 208, 210, 211
Anomaly detection 291
Applications in biology 381
Applications in chemistry 381
Applications in cybersecurity 196, 257, 289–291
Applications of complex numbers 256–257
Applications of cryptography 467–469
Applications of exponential functions 264–270
Applications of Fibonacci 196
Applications of graph theory 380–381
Applications of logic 43–49
Applications of number systems 85–89
Argument 30, 251–254, 289, 307
Aristotle 1
Arithmetic in other number systems 89–95
Arithmetic operations 63
Arithmetic sequences 181–188
Artificial intelligence 46–47
ASCII 88, 462
Associative laws 27–28
Asymmetric cryptography 447–448
Augmented matrix 167–168
Authentication 88, 445, 466–468
Authentication protocols 88
Axioms of probability 347–348

b

Balanced ternary 82–84
Bar graph 297–299
Base-n systems 84
Bernoulli distribution 357
Biconditional 18, 20–22, 49
Binary addition 56–57
Binary arithmetic 56–61
Binary data representation 87
Binary division 59–61
Binary multiplication 59
Binary numbers 52–64
Binary subtraction 57–59
Binary to hexadecimal 78–79
Binary to octal 66–67
Binomial coefficients 199–204, 323, 357, 361
Binomial distribution 323–329, 357–358
Binomial theorem 200–204
Bipartite graphs 405–407
Bipartite matching 444
Bit manipulation 52
Block cipher mode 460, 470
Block ciphers 459, 460
Boolean algebra 43–44
Boolean expression simplification 9–16, 19, 20
Breadth-first search (BFS) 409, 415–418
Brute force attacks 469

Mathematics in Cybersecurity, First Edition. Alfred Basta, Stephan DeLong, and Stavros Basta.
© 2026 John Wiley & Sons, Inc. All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies. Published 2026 by John Wiley & Sons, Inc.
Companion website: www.wiley.com/go/bastamic

c

Caesar cipher 446, 449–451
Caesar shift techniques 449–451
Cardinality 3, 7, 8, 10, 12
Cayley, Arthur 379
Central limit theorem 373–375
Central tendency 304–307
Central tendency measures 304–307
Character frequency 470
Chosen plaintext attacks 470–471
Cipher block chaining (CBC) 470–471
Ciphertext 88, 89, 447, 470
Ciphertext analysis 452, 470
Classical ciphers 448–459
Cluster sampling 295–296
Code breaking 453, 460, 472–475
Color representation 85
Columnar transposition 455–459
Combinational circuit optimization 86
Combinational circuits 86
Combinations 22, 33, 199–204, 352, 354
Common difference 181
Common ratio 189–191
Communication networks 444
Commutative laws 27
Complement laws 44
Complement of a set 14
Complete bipartite graphs 405
Complete graphs 402–405
Complex conjugates 250
Complex number operations 245, 248, 249, 252, 257
Complex numbers 241–257
Complex plane 242–243
Components 6, 402, 420, 422–423, 431
Composite numbers 88, 204, 462
Compound propositions 19–20, 26, 29
Computational complexity 145
Computational graph theory 381
Computer architecture 85–86
Computer memory 52, 75, 78, 85
Conditional probability 350–351
Conditional probability calculation 350–351
Conjunction 18–21, 27, 49, 468
Connected component analysis 419
Connected graphs 379, 386, 413, 418–423, 426, 428, 436, 438, 441
Contingency 24–25, 49
Continuous random variables 355, 368, 372–373
Contradiction 24, 49, 265
Convergence 191
Conversion between systems 209
Correlation coefficient 329–330
Cryptanalysis 446, 448, 469–471, 475
Cryptanalysis techniques 446, 448, 458, 469–471, 475
Cryptographic attacks 469
Cryptographic hash functions 88, 347, 466
Cryptographic key generation 462
Cryptographic protocols 347, 465
Cryptography xix, 84, 88, 158, 161, 196, 203, 204, 257, 289, 347, 408, 445–475
Cycle detection 415

d

Data authentication 468
Data Encryption Standard (DES) 88, 446, 460
Data encryption techniques 196, 461
Data integrity 468
Data integrity verification 88, 466, 468
Data representation 85
De Morgan, Augustus 1
De Morgan's laws 29, 44
Decimal system 51–53, 56, 70, 74, 93, 94
Decimal to binary 76, 81
Decimal to hexadecimal 76–77
Decryption 46, 289, 445, 447, 462
Decryption processes 289
Degree of vertex 381–384, 391, 395, 399, 401, 407
Degrees 7, 207–211, 218, 237, 252, 277, 402
Density of graphs 394
Dependent events 351–352
Depth-first search (DFS) 409, 413
DES encryption 88, 446, 447, 460
Diameter 322, 403
Diameter calculation 322
Difference of sets 10–11
Diffie–Hellman 465
Diffie–Hellman protocol 89, 465
Digital circuit design 18, 33, 52, 86
Digital logic 81, 86–87
Digital logic design 86–87
Digital signatures 468–469
Digital systems 43–44, 87

Digraphs 378, 396–397
Dijkstra's algorithm 379, 380, 398, 410, 433, 434, 436
Directed edges 383, 396–397, 399, 400
Directed graphs 386, 387, 391, 396–397, 420, 423
Discrete data distributions 356–366
Discrete logarithm 465
Discrete mathematics 18, 309
Discrete random variables 354–355, 364–365
Disjoint set structures 438
Disjoint sets 13, 406
Disjunction 18, 20, 21, 27, 49
Disjunctive syllogism 32
Dispersion measures 307–311
Distance calculations 433
Distributive laws 28–29
Division method 54–55, 65
Division of complex numbers 248–251
Double negation 30
Double-angle identities 234–235

e

Edge list 411–412
Edge weights 433
Edges 377, 381–385, 390–392, 394, 396, 399–403, 405, 407, 409, 411, 419, 423, 425, 426, 437
Elements 2–3, 7, 8, 148, 150
Elliptic curve cryptography 257
Empirical probability 346–347
Empty set 5–6, 8
Encryption 46, 88, 196, 289, 445–447, 451, 453–455, 462, 465, 467, 468
Encryption algorithms 45, 88, 196, 257, 447, 448, 460, 470
Encryption key management 196, 468, 470
Encryption processes 289, 453, 454
Entropy in information theory 380
Enumeration 379
Equivalence of sets 12
Erdős, Paul 380
Error detection 36
Euler, Leonhard 379
Eulerian graphs 427–431
Eulerian trails 430
Euler's bridge problem 379
Euler's formula 425–427
Event independent 349–350
Events 293, 341–352
Existential quantifier 41
Expected value 364–365, 372, 452
Expected value calculations 364, 365, 372
Expert systems 46, 47
Exponential decay 259, 260
Exponential distribution 369–372
Exponential equations 183, 284–286
Exponential form 254–257
Exponential functions 259–270
Exponential growth 196, 290

f

Factorial computation 353, 469
Factorial notation 199–204
Fibonacci in cryptography 196
Fibonacci sequence 192–196
File permissions 64, 86
Finite series 184–185
Finite sets 6–7, 347, 354, 464
First-order logic 40, 47
Flip-flops 39
Floyd–Warshall algorithm 419, 434–436
Ford–Fulkerson algorithm 398, 441–444
Frequency analysis 451, 452, 470
Frequency distributions 470

g

Galilei, Galileo 1
Galois fields 464
Gaussian distribution 311, 368
Gaussian elimination 167–176
GCD calculations 462
Generating formulas 178, 179
Geometric distribution 358–360
Geometric interpretation 245–247
Geometric probability 358–360
Geometric sequences 189–191
Graph algorithms 432–444
Graph coloring 379
Graph connectivity 378, 380, 381, 385, 391, 394–397, 403–405, 407, 408, 411, 415, 417–423
Graph theory 377–444
Graph theory applications 380–381
Graph theory history 379–380

Graph traversal algorithms 410, 413–419, 423, 432
Graph traversals 413–418
Greatest common divisor 462
Greedy algorithms 436, 438, 441

h

Hacking 447
Half-adder circuits 86
Half-angle identities 236–240
Hamilton, William 379
Hamiltonian cycle 379, 430–432
Hamiltonian graphs 430–432
Hash collision 347
Hash functions 466–467
Hexadecimal arithmetic 89
Hexadecimal color codes 75, 78
Hexadecimal conversion 77–78, 85
Hexadecimal system 64
Hexadecimal to binary 78–79
Histogram 303
Histogram analysis 303
Historical cryptography 446
HMAC 466–467
HMAC authentication 466–467
Hypergeometric distribution 361–364
Hypothetical syllogism 32

i

Icosian game 379
Idempotent laws 29–30
Identity matrix 149, 158
If and only if 18, 21–22, 224
IF-THEN 20
Imaginary unit 242, 248
Implication 18, 20, 21, 25, 49
Incidence matrix 391, 411
Incidence relationships 411
Incident edges 390–391, 419
Independent events 349–350
Induction 196–198
Inductive hypothesis 198
Inference engine 46, 47
Inference engine operation 46, 47
Inference rules 32
Infinite geometric sequences 189–191
Infinite sets 6–7, 12, 196
Information security 196, 445
Integer sets 7, 10, 11
Intersection 9–10, 15, 124, 125, 128, 215
Intersection operations 9–10
Intrusion detection 291
Intrusion prevention 196, 291
Inverse functions 259
Inverse matrices 161
Inverse trigonometric functions 224–226
Invertible matrices 158–165, 173, 174

k

Key exchange 89, 464, 465
Key exchange protocols 89, 448
Keys 196, 289, 447, 448, 465, 468
Kirchhoff, Gustave 379
Kirchhoff's laws 379
Knowledge base 46, 47
Knowledge representation 47
Known plaintext attacks 470
König, Dénes 379
Königsberg bridge 379, 428
Königsberg problem 427, 428, 430
Kruskal's algorithm 398, 412, 436–438, 441
Kruskal's spanning tree 398, 436, 441

l

Large number factorization 463
Least squares 336, 337, 339, 351
Line graph 300–301
Linear algebra applications 148
Linear correlation 329–340
Linear equations 123–176
Linear transformation 148, 158, 161
Logarithmic equations 284–286
Logarithmic functions 279–283, 289–291
Logarithms 271–281, 286, 464
Logarithms properties 272–275, 286, 288
Logic 1, 17–73, 81, 86–87, 239
Logic circuit design 43, 44
Logic circuits 36–38
Logic gates 33–40
Logic in AI 46, 47
Logical connectives 18, 20
Logical equivalence 26–30
Logical fallacies 47, 48
Logical inference 30–32, 49

Logical reasoning 18, 46–47
Loops 385, 394, 399–402, 409

m

Magnitude 62, 249, 251–256
Magnitude and argument 251–252
Mathematical induction 196–198
Mathematical induction principles 196–198
Mathematical logic 1
Mathematical models 289
Matrices 148–176, 409, 411, 412
Matrix addition 149–153
Matrix determinants 158–165, 173, 174
Matrix equation 165–167
Matrix inverse 158–161, 163, 164
Matrix multiplication 153–158
Matrix operations 149–165
Matrix rank 159, 162, 163
Matrix transpose 159, 163, 455
Max-Flow Min-Cut 441, 443–444
Maximum flow 442
MD5 88
Mean 5, 39, 110, 117, 123, 250, 293–339, 365, 368, 373, 375, 413
Mean and median 304
Measures of dispersion 307–311
Median 304–307, 312
Memory addressing 87
Memory addressing systems 85, 87
Message authentication codes 466
Message digest 88, 466
Minimum cut 398, 409, 411, 444
Minimum spanning tree 398, 436–441
Mixed graphs 399–400
Mode 304, 306, 307, 471
Modular arithmetic 88, 203, 449
Modular exponentiation 88
Modulo operations 204
Modus ponens 31–32
Modus tollens 31–32
Monoalphabetic substitution 451–453
Multigraphs 399–402
Multiple edges 391–394, 399, 400, 402
Multiplication principle 350, 352
Multiplication rule 349, 351–352
Multiplicative inverse 158–165
Mutually exclusive 343, 348, 349

n

NAND gate 34–35
Natural base 263–264
Natural logarithms 272
Negation 18–20, 30, 41–42, 49, 246
Network analysis 409
Network flow 411
Network flow algorithms 441–444
Network reliability 419
Network security 196, 289
Network topology 408, 423
Network traffic 289
Newton, Sir Isaac 1
Nodes 381–383, 386, 408, 419
Non-repudiation 88, 445, 468
NOR gate 35
Normal curve 318, 319, 321, 322
Normal distribution 311–322, 368–369, 374, 375
NOT gate 34, 35, 38
Null set 5–6
Number line 103
Number systems 51–102
Number theory 195
Number theory applications 195, 203
Numerical methods 134, 296

o

Octal conversion 65, 67–68
Octal system 86
Octal to hexadecimal 81
One-way functions 464, 466
Operations on sets 8–11
Optimization algorithms 380, 407, 409, 432
OR gate 33
OR operations 45
Order of operations 23
Ordered pairs 5, 103, 105, 128, 377, 396, 397, 412
Outcomes 323, 341–348

p

Parallel edges 401
Parent–child trees 424
Pascal's triangle 199, 203
Password generation 196
Password management 196
Path algorithms 410
Permutation computation 353–354

Permutations 352, 353
Phi function 192
Pie chart 299–300
Place value 52–54, 65
Place value systems 52
Plaintext 88, 89, 447, 449, 451, 453–455, 461, 470
Plaintext format 88
Planar graph properties 425, 427
Planar graphs 424–427
Point-slope form 110, 111, 113
Poisson distribution 360–361
Polar form 252–254
Polyalphabetic substitution 453
Polynomial equations 288
Population parameters 373
Positional notation 52–54
Positional number systems 65
Post-quantum algorithms 472
Post-quantum cryptography 472
Power set 8
Power set notation 6, 8
Powers of i 248
Predicate logic 40–43
Predicates 40, 49
Prime factorization 274, 462
Prime generation 462
Prime number theorem 41–43, 88–89, 257, 289, 462, 463, 465
Prime numbers 88–89, 257
Prim's algorithm 398, 436, 438, 439, 441
Probability 199, 293, 318, 321, 323, 324, 328, 341–375, 385
Probability axioms 347–348
Probability calculations 199, 347, 348
Probability density 328, 367–368, 372, 373
Probability density functions 367–368
Probability distributions 367–373
Probability mass 355
Probability mass function 355–356
Probability theory 352, 373
Product rule 273
Proof techniques 32
Proper subsets 7, 10, 431
Propositional logic 19–25, 40
Propositions 18–20, 22, 26, 29, 49
Protocol security 467
Pseudo-random number generation 204
Pseudo-random number generators (PRNGs) 204, 295
Public and private keys 257, 448, 461, 464
Public key cryptography 462, 465
Public key systems 88
Pythagorean identities 226, 227, 230, 235–237, 239
Pythagorean theorem 216

q

Quadratic formula 241, 288
Quantifier negation 41–42
Quantifiers 73
Quantum cryptography 446, 471, 472
Quantum key distribution (QKD) 471
Queue structures 195, 418

r

Radians 208–209, 252, 253
Radix notation 51, 84
Rail fence cipher 453–455, 459
Random number generation 204
Random sampling 295
Random variables 354–355, 364–365, 368, 372
Range 77, 100, 101, 154, 222, 226, 281–283, 307–311, 329, 355, 372
Range calculations 307–308, 310, 311, 372
Rational numbers 12
RC4 cipher 460–461
Real numbers 96
Recurrence relations 204
Recursive algorithms 18, 413
Recursive formulas 178–179
Reflexive relations 12
Regression analysis 158
Regression coefficients 337
Regular graphs 407–408
Right triangle trigonometry 214–215
Right triangles 214–215, 217
Risk analysis 290
Rivest–Shamir–Adleman 446
Rivest–Shamir–Adleman algorithm 88, 204, 257, 446, 462
Root finding 217
Roster notation 3
Rounding errors 98
Row operations 140, 159, 168

RSA cryptography 289
RSA cryptosystem 204, 462–463, 465
RSA encryption 88
RSA key generation 88, 89, 204, 257, 289, 446, 462, 465, 470

S

Sample mean 304, 373–375
Sample space 341–343, 345, 348, 355
Sample standard deviation 309
Sample variance 308–311
Sampling methods 293
Sampling techniques 293–296
Scalar multiplication 154
Scalar operations 154
Scatter plot 302
Scatter plot analysis 302–303
Secant function 216
Secure communication 445, 467–468
Secure Hash 466
Security analysis 196, 257, 289
Security protocols 467
Self-loops 409
Semi-Eulerian trails 430
Sequence convergence 191
Sequences 177–204
Sequential circuits 87
Sequential logic 38–40
Series summation 184
Set builder notation 4, 5, 9, 128, 141
Set complement 14
Set membership 3–5
Set notation 3–5
Set operations 8–11
Set relations 11–16
Set theory 6, 8
Sets 1–16, 305, 348, 405, 406
SHA-3 490
SHA-256 347, 466
Shannon, Claude 380
Shannon entropy 380
Shortest path 398, 403, 410, 419, 432, 434
Sigma notation 184
Signal processing 256
Signed binary 61
Significance testing 278, 291, 330
Sign-magnitude 62–63
Simple graphs 398–400
Simple random sampling 295
Simplification law 30
Sine and cosine 227, 228
Single-key systems 460
Singleton elements 6
Singleton sets 6
Skewness 312
Slope calculations 106–108
Social networks 380, 396, 397, 399, 402, 404, 421
Soundness 54
Spanning tree algorithms 420, 436, 441
Square matrices 158, 387
Square matrix 149, 158, 159, 408
Square root 217, 242
Stack structures 415
Standard deviation 309–311, 318, 321, 373, 375
Standard normal 312–322, 374, 375
State diagrams 87
State tables 87
State transitions 87
Statistical graphs 297–303
Statistics 293–339, 341, 373
Stratified sampling 293, 296–297
Stream cipher operation 461
Stream ciphers 484, 485
Strongly connected graphs 420–422
Subgraph 397, 403, 415, 417–420, 422, 423, 431
Subset 7, 199, 431, 442, l8
Substitution ciphers 448–453
Substitution methods 128–133, 138, 139
Subtraction method 55
Subtraction of complex numbers 243–248
Sum and difference identities 222, 230–232
Summation formulas 77, 78
Summation notation 177–204
Superset 7
Syllogism 32
Symbolic logic 31
Symmetric cryptography 447–448
Symmetric encryption 465
Symmetric key systems 445, 446, 459–461, 464
Symmetric relations 384
Systematic sampling 294–295
Systems of linear equations 123–176

t

Tangent function 223, 229, 252
Tautology 24, 49
Ternary logic 81
Ternary system 83, 84
Theorem 1, 199–204, 216, 373–375
Theoretical foundations xix
Theoretical probability 344
Threat analysis 196
Time complexity 410, 412, 415, 417, 421, 434, 436, 438, 441, 443
Topological sorting 415
Traffic analysis 196, 289–290
Traffic patterns 291
Transformation matrices 154, 158
Transitive relations 12
Transposition ciphers 453
Tree algorithms 436
Tree traversal 413, 418
Trees 379, 423
Trigonometric equations 226
Trigonometric functions 214–217, 219–226, 230–240
Trigonometric identities 226–230
Trigonometric ratios 216
Truth tables 22, 37
Truth values 20, 22–24
Turing, Alan 1, 453
Two's complement 64
Two's complement representation 61, 63, 64
Types of graphs 395–408
Types of logic gates 34
Types of sets 5–8

u

Unconditional security 471
Undirected edges 383–384, 399
Undirected graphs 395, 396
Unicode 85
Unicode encoding 85
Uniform distribution 368
Union of sets 8–9
Union operations 8, 9
Unit circle 237
Universal quantifier 40
Universal sets 13
Upper bound 184

v

Valid arguments 17, 48
Validity 17, 30, 32, 286
Variable domains 40, 41, 43, 367
Variance 307–311, 364, 365, 372
Variance calculations 308–311, 365, 372, 373
Vector addition 245
Vector operations 158
Vectors 245, 257
Venn diagram construction 15
Venn diagrams 14–16
Verification methods 29
Vertex coloring 406
Vertex connectivity 403, 415, 420, 421, 423
Vertex degree 381–384, 391, 395, 397, 399, 401, 407
Vertical asymptotes 224, 279, 280
Vertices 377–388, 391, 395, 397–413, 418–420, 422–424, 430–434, 438
Vulnerability analysis 290
Vulnerability assessment 290

w

Weakly connected components 422
Weakly connected graphs 421–422
Weighted graphs 432
Whole numbers 3, 7, 11, 24, 97, 98, 210, 216, 241, 294, 313, 348, 385
Wittgenstein, Ludwig 1
Worst-case analysis 412

x

XNOR gate 36
XOR cipher 470
XOR encryption 45
XOR gate 35–37
XOR operations 46, 196, 460, 461

z

Zero divisors 60, 61
Zero matrix 149
Zero polynomial 288
Z-score 313, 316, 317

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.