ECE 606, Algorithms

Mahesh Tripunitara
tripunit@uwaterloo.ca
ECE, University of Waterloo

Acknowledgements

Considerable material is straight out of CLRS, T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms," MIT Press, Editions II and III. Some material is from the following.

- J. Thistle and M. Tripunitara, ECE 108 "textbook."
- W. Conradie and V. Goranko, "Logic and Discrete Mathematics, a Concise Introduction," Wiley.
- G. P. Hochschild, "Perspectives of Elementary Mathematics," Springer-Verlag.
- P. J. Cameron, "Sets, Logic and Categories," Springer.
- M. Huth and M. Ryan, "Logic in Computer Science, Modelling and Reasoning about Systems," Cambridge.
- S. Dasgupta, C. Papadimitriou and U. Vazirani, "Algorithms," McGraw-Hill.
- Elyse Yeager, `http://www.math.ubc.ca/~elyse/220/2016/9Disproof.pdf`
- J. Kleinberg and E. Tardos, "Algorithm Design," Addison Wesley Longman.
- S. Arora and B. Barak, "Computational Complexity – a Modern Approach," Cambridge.
- J. Hopcroft, R. Motwani and J. Ullman, "Introduction to Automata Theory, Languages, and Computation," Addison Wesley.
- V. Vazirani, "Approximation Algorithms," Springer.

Contents

Lecture 1

- Discrete math review.
- Introduction to Python 3.

Discrete math is a collection of branches of mathematics that deals with discrete, as opposed to continuous, structures. An example of a discrete structure is the set of integers, $\{\ldots, -2, -1, 0, 1, \ldots\}$. We call those "discrete" because they are a collection of "distinct and unconnected elements," as defined in the Merriam-Webster's dictionary. The real numbers, on the other hand, are not discrete: between any two real numbers, we can find another real number.

Logic is a set of principles for systematic reasoning. For example, if I know that Arthur is a cat, and that every cat is a carnivore, then I can infer that Arthur is a carnivore.

Discrete math and logic are useful for understanding and analyzing algorithms. We review the following here.

- Sets and functions.
- Propositional logic, and the universal and existential quantifers.
- Techniques for proving an assertion or its negation.
- Discrete probability and expectation.

6

Sets, relations and functions

From Cormen, et al., “Introduction to Algorithms.”

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

B.1 Sets

A ***set*** is a collection of distinguishable objects, called its ***members*** or ***elements***. If an object x is a member of a set S, we write $x \in S$ (read "x is a member of S" or, more briefly, "x is in S"). If x is not a member of S, we write $x \notin S$. We can describe a set by explicitly listing its members as a list inside braces. For example, we can define a set S to contain precisely the numbers 1, 2, and 3 by writing $S = \{1, 2, 3\}$. Since 2 is a member of the set S, we can write $2 \in S$, and since 4 is not a member, we have $4 \notin S$. A set cannot contain the same object more than once,[1] and its elements are not ordered. Two sets A and B are ***equal***, written $A = B$, if they contain the same elements. For example, $\{1, 2, 3, 1\} = \{1, 2, 3\} = \{3, 2, 1\}$.

We adopt special notations for frequently encountered sets.

- $\emptyset$ denotes the ***empty set***, that is, the set containing no members.
- **Z** denotes the set of ***integers***, that is, the set $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
- **R** denotes the set of ***real numbers***.
- **N** denotes the set of ***natural numbers***, that is, the set $\{0, 1, 2, \ldots\}$.[2]

[1] A variation of a set, which can contain the same object more than once, is called a ***multiset***.

[2] Some authors start the natural numbers with 1 instead of 0. The modern trend seems to be to start with 0.

If all the elements of a set A are contained in a set B, that is, if $x \in A$ implies $x \in B$, then we write $A \subseteq B$ and say that A is a ***subset*** of B. A set A is a ***proper subset*** of B, written $A \subset B$, if $A \subseteq B$ but $A \neq B$. (Some authors use the symbol "$\subset$" to denote the ordinary subset relation, rather than the proper-subset relation.) For any set A, we have $A \subseteq A$. For two sets A and B, we have $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$. For any three sets A, B, and C, if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$. For any set A, we have $\emptyset \subseteq A$.

We sometimes define sets in terms of other sets. Given a set A, we can define a set $B \subseteq A$ by stating a property that distinguishes the elements of B. For example, we can define the set of even integers by $\{x : x \in \mathbf{Z} \text{ and } x/2 \text{ is an integer}\}$. The colon in this notation is read "such that." (Some authors use a vertical bar in place of the colon.)

Given two sets A and B, we can also define new sets by applying ***set operations***:

- The ***intersection*** of sets A and B is the set

  $$A \cap B = \{x : x \in A \text{ and } x \in B\} .$$

- The ***union*** of sets A and B is the set

  $$A \cup B = \{x : x \in A \text{ or } x \in B\} .$$

- The ***difference*** between two sets A and B is the set

  $$A - B = \{x : x \in A \text{ and } x \notin B\} .$$

Set operations obey the following laws.

Empty set laws:

$$\begin{aligned} A \cap \emptyset &= \emptyset , \\ A \cup \emptyset &= A . \end{aligned}$$

Idempotency laws:

$$\begin{aligned} A \cap A &= A , \\ A \cup A &= A . \end{aligned}$$

Commutative laws:

$$\begin{aligned} A \cap B &= B \cap A , \\ A \cup B &= B \cup A . \end{aligned}$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

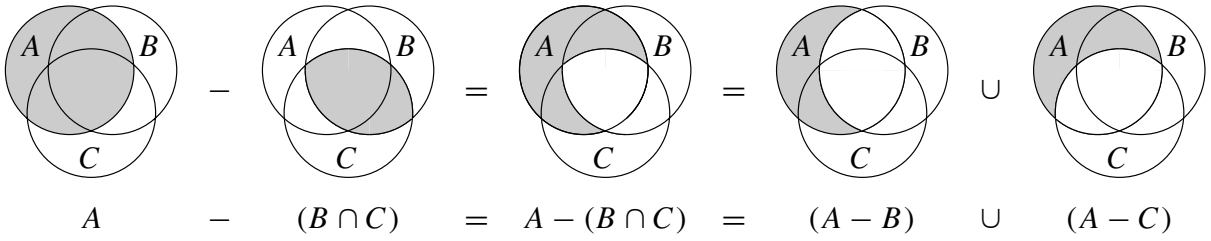

Figure B.1 A Venn diagram illustrating the first of DeMorgan's laws (B.2). Each of the sets A, B, and C is represented as a circle.

Associative laws:

$$\begin{aligned} A \cap (B \cap C) &= (A \cap B) \cap C \,, \\ A \cup (B \cup C) &= (A \cup B) \cup C \,. \end{aligned}$$

Distributive laws:

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \,, \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \,. \end{aligned} \tag{B.1}$$

Absorption laws:

$$\begin{aligned} A \cap (A \cup B) &= A \,, \\ A \cup (A \cap B) &= A \,. \end{aligned}$$

DeMorgan's laws:

$$\begin{aligned} A - (B \cap C) &= (A - B) \cup (A - C) \,, \\ A - (B \cup C) &= (A - B) \cap (A - C) \,. \end{aligned} \tag{B.2}$$

The first of DeMorgan's laws is illustrated in Figure B.1, using a ***Venn diagram***, a graphical picture in which sets are represented as regions of the plane.

Often, all the sets under consideration are subsets of some larger set U called the ***universe***. For example, if we are considering various sets made up only of integers, the set $\mathbf{Z}$ of integers is an appropriate universe. Given a universe U, we define the ***complement*** of a set A as $\overline{A} = U - A$. For any set $A \subseteq U$, we have the following laws:

$$\begin{aligned} \overline{\overline{A}} &= A \,, \\ A \cap \overline{A} &= \emptyset \,, \\ A \cup \overline{A} &= U \,. \end{aligned}$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

DeMorgan's laws (B.2) can be rewritten with complements. For any two sets $B, C \subseteq U$, we have

$$\begin{aligned} \overline{B \cap C} &= \overline{B} \cup \overline{C} , \\ \overline{B \cup C} &= \overline{B} \cap \overline{C} . \end{aligned}$$

Two sets A and B are ***disjoint*** if they have no elements in common, that is, if $A \cap B = \emptyset$. A collection $\mathcal{S} = \{S_i\}$ of nonempty sets forms a ***partition*** of a set S if

- the sets are ***pairwise disjoint***, that is, $S_i, S_j \in \mathcal{S}$ and $i \neq j$ imply $S_i \cap S_j = \emptyset$, and
- their union is S, that is,

$$S = \bigcup_{S_i \in \mathcal{S}} S_i .$$

In other words, $\mathcal{S}$ forms a partition of S if each element of S appears in exactly one $S_i \in \mathcal{S}$.

The number of elements in a set is called the ***cardinality*** (or ***size***) of the set, denoted $|S|$. Two sets have the same cardinality if their elements can be put into a one-to-one correspondence. The cardinality of the empty set is $|\emptyset| = 0$. If the cardinality of a set is a natural number, we say the set is ***finite***; otherwise, it is ***infinite***. An infinite set that can be put into a one-to-one correspondence with the natural numbers $\mathbf{N}$ is ***countably infinite***; otherwise, it is ***uncountable***. The integers $\mathbf{Z}$ are countable, but the reals $\mathbf{R}$ are uncountable.

For any two finite sets A and B, we have the identity

$$|A \cup B| = |A| + |B| - |A \cap B| , \tag{B.3}$$

from which we can conclude that

$$|A \cup B| \leq |A| + |B| .$$

If A and B are disjoint, then $|A \cap B| = 0$ and thus $|A \cup B| = |A| + |B|$. If $A \subseteq B$, then $|A| \leq |B|$.

A finite set of n elements is sometimes called an ***n-set***. A 1-set is called a ***singleton***. A subset of k elements of a set is sometimes called a ***k-subset***.

The set of all subsets of a set S, including the empty set and S itself, is denoted 2^S and is called the ***power set*** of S. For example, $2^{\{a,b\}} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$. The power set of a finite set S has cardinality $2^{|S|}$.

We sometimes care about setlike structures in which the elements are ordered. An ***ordered pair*** of two elements a and b is denoted (a, b) and can be defined formally as the set $(a, b) = \{a, \{a, b\}\}$. Thus, the ordered pair (a, b) is *not* the same as the ordered pair (b, a).

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

The ***Cartesian product*** of two sets A and B, denoted $A \times B$, is the set of all ordered pairs such that the first element of the pair is an element of A and the second is an element of B. More formally,

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\} \ .$$

For example, $\{a, b\} \times \{a, b, c\} = \{(a, a), (a, b), (a, c), (b, a), (b, b), (b, c)\}$. When A and B are finite sets, the cardinality of their Cartesian product is

$$|A \times B| = |A| \cdot |B| \ . \tag{B.4}$$

The Cartesian product of n sets $A_1, A_2, \ldots, A_n$ is the set of ***n-tuples***

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

whose cardinality is

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \cdot |A_2| \cdots |A_n|$$

if all sets are finite. We denote an n-fold Cartesian product over a single set A by the set

$$A^n = A \times A \times \cdots \times A \ ,$$

whose cardinality is $|A^n| = |A|^n$ if A is finite. An n-tuple can also be viewed as a finite sequence of length n (see page 1078).

Copyright @ 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

B.2 Relations

A ***binary relation*** R on two sets A and B is a subset of the Cartesian product $A \times B$. If $(a, b) \in R$, we sometimes write $a\ R\ b$. When we say that R is a binary relation on a set A, we mean that R is a subset of $A \times A$. For example, the "less than" relation on the natural numbers is the set $\{(a, b) : a, b \in \mathbf{N} \text{ and } a < b\}$. An n-ary relation on sets $A_1, A_2, \ldots, A_n$ is a subset of $A_1 \times A_2 \times \cdots \times A_n$.

A binary relation $R \subseteq A \times A$ is ***reflexive*** if

$$a\ R\ a$$

for all $a \in A$. For example, "$=$" and "$\leq$" are reflexive relations on $\mathbf{N}$, but "$<$" is not. The relation R is ***symmetric*** if

$$a\ R\ b \text{ implies } b\ R\ a$$

for all $a, b \in A$. For example, "$=$" is symmetric, but "$<$" and "$\leq$" are not. The relation R is ***transitive*** if

$$a\ R\ b \text{ and } b\ R\ c \text{ imply } a\ R\ c$$

for all $a, b, c \in A$. For example, the relations "$<$," "$\leq$," and "$=$" are transitive, but the relation $R = \{(a, b) : a, b \in \mathbf{N} \text{ and } a = b - 1\}$ is not, since $3\ R\ 4$ and $4\ R\ 5$ do not imply $3\ R\ 5$.

A relation that is reflexive, symmetric, and transitive is an ***equivalence relation***. For example, "$=$" is an equivalence relation on the natural numbers, but "$<$" is not. If R is an equivalence relation on a set A, then for $a \in A$, the ***equivalence class*** of a is the set $[a] = \{b \in A : a\ R\ b\}$, that is, the set of all elements equivalent to a. For example, if we define $R = \{(a, b) : a, b \in \mathbf{N} \text{ and } a + b \text{ is an even number}\}$, then R is an equivalence relation, since $a + a$ is even (reflexive), $a + b$ is even implies $b + a$ is even (symmetric), and $a + b$ is even and $b + c$ is even imply $a + c$ is even (transitive). The equivalence class of 4 is $[4] = \{0, 2, 4, 6, \ldots\}$, and the

equivalence class of 3 is $[3] = \{1, 3, 5, 7, \ldots\}$. A basic theorem of equivalence classes is the following.

Theorem B.1 (An equivalence relation is the same as a partition)
The equivalence classes of any equivalence relation R on a set A form a partition of A, and any partition of A determines an equivalence relation on A for which the sets in the partition are the equivalence classes.

Proof For the first part of the proof, we must show that the equivalence classes of R are nonempty, pairwise-disjoint sets whose union is A. Because R is reflexive, $a \in [a]$, and so the equivalence classes are nonempty; moreover, since every element $a \in A$ belongs to the equivalence class $[a]$, the union of the equivalence classes is A. It remains to show that the equivalence classes are pairwise disjoint, that is, if two equivalence classes $[a]$ and $[b]$ have an element c in common, then they are in fact the same set. Now $a\,R\,c$ and $b\,R\,c$, which by symmetry and transitivity imply $a\,R\,b$. Thus, for any arbitrary element $x \in [a]$, we have $x\,R\,a$ implies $x\,R\,b$, and thus $[a] \subseteq [b]$. Similarly, $[b] \subseteq [a]$, and thus $[a] = [b]$.

For the second part of the proof, let $\mathcal{A} = \{A_i\}$ be a partition of A, and define $R = \{(a, b) : \text{there exists } i \text{ such that } a \in A_i \text{ and } b \in A_i\}$. We claim that R is an equivalence relation on A. Reflexivity holds, since $a \in A_i$ implies $a\,R\,a$. Symmetry holds, because if $a\,R\,b$, then a and b are in the same set A_i, and hence $b\,R\,a$. If $a\,R\,b$ and $b\,R\,c$, then all three elements are in the same set, and thus $a\,R\,c$ and transitivity holds. To see that the sets in the partition are the equivalence classes of R, observe that if $a \in A_i$, then $x \in [a]$ implies $x \in A_i$, and $x \in A_i$ implies $x \in [a]$. ■

A binary relation R on a set A is ***antisymmetric*** if

$$a\,R\,b \text{ and } b\,R\,a \text{ imply } a = b\ .$$

For example, the "$\leq$" relation on the natural numbers is antisymmetric, since $a \leq b$ and $b \leq a$ imply $a = b$. A relation that is reflexive, antisymmetric, and transitive is a ***partial order***, and we call a set on which a partial order is defined a ***partially ordered set***. For example, the relation "is a descendant of" is a partial order on the set of all people (if we view individuals as being their own descendants).

In a partially ordered set A, there may be no single "maximum" element a such that $b\,R\,a$ for all $b \in A$. Instead, there may be several ***maximal*** elements a such that for no $b \in A$, where $b \neq a$, is it the case that $a\,R\,b$. For example, in a collection of different-sized boxes there may be several maximal boxes that don't fit inside any other box, yet no single "maximum" box into which any other box will fit.[3]

[3]To be precise, in order for the "fit inside" relation to be a partial order, we need to view a box as fitting inside itself.

Copyright @ 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSCO : eBook Collection (EBSCOhost) - printed on 8/4/2019 9:14 PM via UNIV OF WATERLOO
AN: 139815 ; Cormen, Thomas H..; Introduction to Algorithms
Account: s9860349.main.ehost

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

B.3 Functions

Given two sets A and B, a ***function*** f is a binary relation on $A \times B$ such that for all $a \in A$, there exists precisely one $b \in B$ such that $(a, b) \in f$. The set A is called the ***domain*** of f, and the set B is called the ***codomain*** of f. We sometimes write

$f : A \to B$; and if $(a, b) \in f$, we write $b = f(a)$, since b is uniquely determined by the choice of a.

Intuitively, the function f assigns an element of B to each element of A. No element of A is assigned two different elements of B, but the same element of B can be assigned to two different elements of A. For example, the binary relation

$$f = \{(a, b) : a, b \in \mathbf{N} \text{ and } b = a \bmod 2\}$$

is a function $f : \mathbf{N} \to \{0, 1\}$, since for each natural number a, there is exactly one value b in $\{0, 1\}$ such that $b = a \bmod 2$. For this example, $0 = f(0)$, $1 = f(1)$, $0 = f(2)$, etc. In contrast, the binary relation

$$g = \{(a, b) : a, b \in \mathbf{N} \text{ and } a + b \text{ is even}\}$$

is not a function, since $(1, 3)$ and $(1, 5)$ are both in g, and thus for the choice $a = 1$, there is not precisely one b such that $(a, b) \in g$.

Given a function $f : A \to B$, if $b = f(a)$, we say that a is the ***argument*** of f and that b is the ***value*** of f at a. We can define a function by stating its value for every element of its domain. For example, we might define $f(n) = 2n$ for $n \in \mathbf{N}$, which means $f = \{(n, 2n) : n \in \mathbf{N}\}$. Two functions f and g are ***equal*** if they have the same domain and codomain and if, for all a in the domain, $f(a) = g(a)$.

A ***finite sequence*** of length n is a function f whose domain is the set of n integers $\{0, 1, \ldots, n - 1\}$. We often denote a finite sequence by listing its values: $\langle f(0), f(1), \ldots, f(n - 1)\rangle$. An ***infinite sequence*** is a function whose domain is the set $\mathbf{N}$ of natural numbers. For example, the Fibonacci sequence, defined by recurrence (3.21), is the infinite sequence $\langle 0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots\rangle$.

When the domain of a function f is a Cartesian product, we often omit the extra parentheses surrounding the argument of f. For example, if we had a function $f : A_1 \times A_2 \times \cdots \times A_n \to B$, we would write $b = f(a_1, a_2, \ldots, a_n)$ instead of $b = f((a_1, a_2, \ldots, a_n))$. We also call each a_i an ***argument*** to the function f, though technically the (single) argument to f is the n-tuple $(a_1, a_2, \ldots, a_n)$.

If $f : A \to B$ is a function and $b = f(a)$, then we sometimes say that b is the ***image*** of a under f. The image of a set $A' \subseteq A$ under f is defined by

$$f(A') = \{b \in B : b = f(a) \text{ for some } a \in A'\} .$$

The ***range*** of f is the image of its domain, that is, $f(A)$. For example, the range of the function $f : \mathbf{N} \to \mathbf{N}$ defined by $f(n) = 2n$ is $f(\mathbf{N}) = \{m : m = 2n \text{ for some } n \in \mathbf{N}\}$.

A function is a ***surjection*** if its range is its codomain. For example, the function $f(n) = \lfloor n/2 \rfloor$ is a surjective function from $\mathbf{N}$ to $\mathbf{N}$, since every element in $\mathbf{N}$ appears as the value of f for some argument. In contrast, the function $f(n) = 2n$ is not a surjective function from $\mathbf{N}$ to $\mathbf{N}$, since no argument to f can produce 3 as a value. The function $f(n) = 2n$ is, however, a surjective function from the natural

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

numbers to the even numbers. A surjection $f : A \to B$ is sometimes described as mapping A ***onto*** B. When we say that f is onto, we mean that it is surjective.

A function $f : A \to B$ is an ***injection*** if distinct arguments to f produce distinct values, that is, if $a \neq a'$ implies $f(a) \neq f(a')$. For example, the function $f(n) = 2n$ is an injective function from $\mathbf{N}$ to $\mathbf{N}$, since each even number b is the image under f of at most one element of the domain, namely $b/2$. The function $f(n) = \lfloor n/2 \rfloor$ is not injective, since the value 1 is produced by two arguments: 2 and 3. An injection is sometimes called a ***one-to-one*** function.

A function $f : A \to B$ is a ***bijection*** if it is injective and surjective. For example, the function $f(n) = (-1)^n \lceil n/2 \rceil$ is a bijection from $\mathbf{N}$ to $\mathbf{Z}$:

$$\begin{aligned} 0 &\to 0 , \\ 1 &\to -1 , \\ 2 &\to 1 , \\ 3 &\to -2 , \\ 4 &\to 2 , \\ &\vdots \end{aligned}$$

The function is injective, since no element of $\mathbf{Z}$ is the image of more than one element of $\mathbf{N}$. It is surjective, since every element of $\mathbf{Z}$ appears as the image of some element of $\mathbf{N}$. Hence, the function is bijective. A bijection is sometimes called a ***one-to-one correspondence***, since it pairs elements in the domain and codomain. A bijection from a set A to itself is sometimes called a ***permutation***.

When a function f is bijective, its ***inverse*** f^{-1} is defined as

$$f^{-1}(b) = a \text{ if and only if } f(a) = b .$$

For example, the inverse of the function $f(n) = (-1)^n \lceil n/2 \rceil$ is

$$f^{-1}(m) = \begin{cases} 2m & \text{if } m \geq 0 , \\ -2m - 1 & \text{if } m < 0 . \end{cases}$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Propositional logic, $\forall$ and $\exists$

We now discuss some logic, mostly propositional logic. We then augment it with the universal and existential quantifiers.

Definition 1 (Proposition)**.** *A proposition is a statement with which we are able to associate* true *or* false*. Each of* true *and* false *is called a truth value.*

Examples of propositions:

1. "The Earth is flat."
2. "Not all birds can fly."
3. "A dog is a mammal, and not a bird."

Of course, in the above, Proposition (1) happens to be false, and Propositions (2) and (3) are true. In this context, we should make the rather important observation that in assessing the truth value of a proposition, e.g., "the Earth is flat," we do so in the context of a domain of discourse. For example, if the domain of discourse is a sci-fi novel in which there is an entity called Earth which the novel specifies to be flat, then the truth value of the proposition "the Earth is flat" is true in that domain of discourse. Usually, the domain of discourse is clear from context. For example, when we said confidently that Proposition (1) above is false, we were presumably adopting a domain of discourse in which "Earth" refers to the planet on which we reside.

Examples of statements that are not propositions:

1. "Hey, you!"
2. "Which way is the hotel?"
3. "This statement is false."
4. "The variable x is non-negative."

The first of the above is an exclamation, and the second is a question. As for the third, if it is true, then it is false, and if it is false, then it is true. Thus, we are able to associate neither true nor false with that statement. The fourth refers to a variable that can take on one of several values. Without knowledge of exactly what value x takes at a given moment, we cannot assess the truthfulness of the statement.

In this context, it is interesting and fun to address an old riddle. Suppose one is faced with two persons, call them Alice and Bob, one of whom always speaks the truth, and the other of whom always lies. What questions, when asked of Alice and/or Bob, would reveal which one amongst them is the truth-teller, and which one is the liar?

Suppose we ask one of them, say Alice, whether the other, Bob, would say 'yes' if asked whether Alice is the liar. If Alice is the truth-teller, then she would say 'yes,' because Bob is the liar, and he would answer 'yes' to our question to him, when the correct answer is 'no.' If Alice is the liar, then she would say 'no,' because Bob, as the truth-teller, would say 'yes' if we asked him whether Alice is the liar, and because Alice always lies, she would negate that expected response from Bob.

While devising the right question to ask above certainly takes creativity, underlying the entire exercise is careful logical reasoning. Communicating and inculcating this is exactly our intent with our discussions on propositional logic.

To develop an understanding of propositional logic, we will often deal with propositions abstractly. Specifically, we will adopt usages such as: "Assume that p is a proposition." When we say that, we do not know exactly what the proposition p is. All we know is that p is either true or false.

Given propositions, we can compose them in certain ways to yield other propositions. Some refer to such a new proposition as a *compound* proposition. A proposition that is not compound is called an *atomic* proposition.

The third example of a proposition above, "A dog is a mammal, and not a bird" is an example of a compound proposition. As another example, consider the following two propositions: (i) "The glass is not empty." (ii) "The glass is not full." We can compose them and say, (iii) "The glass is neither empty nor full." Given such a compound proposition, it is necessary to clarify its

semantics, that is, what the truth value of the compound proposition (iii) is as a function of the truth values of its constituent, atomic propositions.

To clarify what we mean, suppose the glass is indeed empty. Then Proposition (i) above is false. This implies that Proposition (iii) is false as well. Similarly, suppose Proposition (iii) is false. Then at least one of Proposition (i) and (ii) is false. A customary way, in propositional logic, to specify a semantics for a proposition that is composed of other propositions is to specify a *truth table*. For our example of Propositions (i)–(iii) above, such a truth table may look like the following.

If "the glass is not empty" is	and "the glass is not full" is	then "the glass is neither empty nor full" is
true	true	true
true	false	false
false	true	false
false	false	false

An important aspect of logic is to carefully distinguish *syntax* from semantics. Syntax refers to the way we write things down. Semantics refers to what they mean. We now specify a syntax for compound propositions. We then clarify what the semantics of each is, via truth tables. The manner in which we specify a syntax for compound propositions is by introducing logical *connectives*, and then asserting that the use of such connectives in particular ways is syntactically valid.

Logical connectives – syntax Given that each of p and q is a proposition, so are the following:

- (p): parenthesization – used to force precedence.
- $\neg p$: negation.
- $p \wedge q$: conjunction.
- $p \vee q$: disjunction.
- $p \implies q$: implication.

- $p \impliedby q$: inference.
- $p \iff q$: if and only if.

Given the above syntax for the use of logical connectives to make new propositions, we can further propose rules via which even more propositions can be derived. They would be similar to the axioms of boolean algebra, which we encounter in the context of digital circuits. We present an example here, but leave more for a future course. For this course, we focus on employing semantics, which we specify using truth tables, to infer more propositions. Similarly, in the context of digital circuits, we usually employ "truth tables," like those employed here, rather than proofs based on the axioms of boolean algebra.

We point out that more connectives can be introduced, for example, $\oplus$, "exclusive-or." It turns out that in propositional logic, the connectives $\neg, \vee$ and $\wedge$ suffice, and all other connectives can be defined using those three only. Also, not all of the connectivities are necessary, in the sense that given a smaller set of them, we can realize the others. In particular, $(\cdot), \neg, \vee$ and $\wedge$ suffice. We introduced $\implies$, $\impliedby$ and $\iff$ as well because those are used heavily in this course for proofs. Consequently, it is useful to directly specify and understand those connectives as well. Similarly, XOR gates are convenient in the context of digital circuits to have, even though its functionality can be realized from NOT, AND and OR gates only.

As an example of the use of purely syntactic derivation, see `proofwiki.org/wiki/Rule_of_Material_Implication/Formulation_1/Forward_Implication/Proof`, which shows a derivation from $p \implies q$ to $\neg p \vee q$.

Logical connectives – semantics The following truth tables are customarily associated with the above propositions that are formed using logical connectives. A truth table specifies, for every possibility of a truth value for the constituent propositions, what the truth value of a compound proposition is. We use T for true, and F for false.

- parenthesization:

p	(p)
T	T
F	F

The above truth table merely emphasizes that the truth value of p is unaffected by parenthesization.

- negation:

p	$\neg p$
T	F
F	T

Example: suppose "the Sun is hot" is true. Then, "the Sun is not hot" is false. The second statement is the manner in which we customarily write the negation of "the Sun is hot" in English.

- conjunction:

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

Example: suppose "the Moon is made of cheese" is false, and "the Sun is hot" is true. Then, "the Moon is made of cheese and the Sun is hot" is false.

- disjunction:

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

Example: suppose "the Moon is made of cheese" is false, and "the Sun is hot" is true. Then, "either the Moon is made of cheese, or the Sun is hot, or both" is true.

- implication:

p	q	$p \implies q$
T	T	T
T	F	F
F	T	T
F	F	T

Example: suppose "the Moon is made of cheese" is false, and "the Sun is hot" is true. Then:

  - "If the Sun is hot, then the Moon is made of cheese" is false.
  - "If the Moon is made of cheese, then the Sun is hot" is true.

– "If the Sun is not hot, then the Moon is made of cheese" is true.

The last two examples illustrate that, in propositional logic, "if p then q" may have a very different meaning than in natural language. In English, it is often used, for instance, to imply a causal relationship between p and q. But given a premise p that is false – for example, "the Sun is not hot" – the implication $p \implies q$ is true for any q, even a completely unrelated proposition q such as "the Moon is made of cheese." So the current truth of $p \implies q$ does not mean that, when the Sun eventually cools, the Moon will then be composed entirely of fermented curd; rather, when the Sun cools, the implication itself will be false: in our truth-functional semantics, the truth value of the compound proposition reflects only the specific truth values of the constituent propositions, and no more profound relationship between those constituent propositions. It may be helpful to think of "if p then q" as shorthand for, "(in any row of the truth table in which $p \implies q$ is true), if p is true, then q is true."

In mathematics, because we use these same truth-functional semantics, if p is false, we say that $p \implies q$ is *vacuously true*, to mean that the implication is true simply by virtue of the falsity of its premise. For example, if p is "x is an element of the empty set," and q is "x has property Q," then $p \implies q$ is (vacuously) true, whatever the property Q: the elements of the empty set can be said to have any property that you like, because there are no such elements.

It is not necessary to read $p \implies q$ as "if p then q"; another common way is to say "p only if q." Again, the proper interpretation is truth-functional. In other words, in our truth-functional semantics, the following two statements are completely equivalent:

– If the Sun is hot, then the Moon is made of cheese.
– The Sun is hot only if the Moon is made of cheese.

- inference:

p	q	$p \impliedby q$
T	T	T
T	F	T
F	T	F
F	F	T

Here the compound proposition is a different way of writing $q \implies p$. It is commonly read, "p if q," but should be interpreted only truth-functionally, and not as implying some deeper relationship between p and q.

Example: suppose "the Moon is made of cheese" is false, and "the Sun is hot" is true. Then:

- "the Sun is hot if the Moon is made of cheese" is true.
- "the Moon is made of cheese if the Sun is hot" is false.
- "the Moon is made of cheese if the Sun is not hot" is true.

- if and only if:

p	q	$p \iff q$
T	T	T
T	F	F
F	T	F
F	F	T

Example: suppose "the Moon is made of cheese" is false, and "the Sun is hot" is true. Then:

- "The Sun is hot if and only if the Moon is made of cheese" is false.
- "The Moon is made of cheese if and only if the Sun is not hot" is true.

Given the above semantics via truth tables, we can now infer several more propositions.

Claim 1. $(p \implies q) \iff (\neg p \vee q)$.

Proof. By truth-table.

p	q	$\neg p$	$p \implies q$	$\neg p \vee q$	$(p \implies q) \iff (\neg p \vee q)$
F	F	T	T	T	T
F	T	T	T	T	T
T	F	F	F	F	T
T	T	F	T	T	T

□

We claim that the above is a valid proof for the claim because for every possible combination of truth values for p and q, we have shown that the proposition in the claim is true. We now make and prove two more claims. The first, which is an implication, has a special name, and is useful for carrying out some proofs. Given $p \implies q$, we call the proposition $\neg q \implies \neg p$ its *contrapositive*. The contrapositive of an implication is different from the *converse*: the converse of $p \implies q$ is $q \implies p$. It turns out that $(p \implies q) \iff (\neg q \implies \neg p)$, that is, an implication and its contrapositive are completely equivalent from the standpoint of their respective truth values. However, given a proposition $p \implies q$, its converse, $q \implies p$, is not necessarily true.

For example, suppose you know that if it rains, then I carry an umbrella. You happen to observe that I am carrying an umbrella. Can you infer anything, for example, that it is raining? The answer is no, not necessarily. On the other hand, suppose you observe that I am not carrying an umbrella. Can you infer anything? The answer is yes, you can infer that it is not raining.

Claim 2. $(p \implies q) \iff (\neg q \implies \neg p)$.

Proof. We prove by truth table.

p	q	$\neg p$	$\neg q$	$p \implies q$	$\neg q \implies \neg p$	$(p \implies q) \iff (\neg q \implies \neg p)$
F	F	T	T	T	T	T
F	T	T	F	T	T	T
T	F	F	T	F	F	T
T	T	F	F	T	T	T

□

We now assert something that is perhaps not as easy to prove. If only because it involves three propositions, p, q and r. But again, careful use of the truth table enables us to carry out the proof somewhat mechanically.

Claim 3. $(p \implies q) \implies (p \vee r \implies q \vee r)$.

Proof. By truth table.

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

□

Perhaps the trickiest part of the truth table in the above proof is intuiting the truth value of the last column when $p \implies q$ is false. Recall that the proposition $\phi \implies \psi$ is true whenever ϕ is false. And in this case, ϕ is $p \implies q$.

A number of other useful propositions can similarly be inferred from the truth tables. Following are some useful propositions, and names we associate with them when perceived as properties.

- $(p \vee q) \iff (q \vee p)$ – commutativity of $\vee$.
- $(p \wedge q) \iff (q \wedge p)$ – commutativity of $\wedge$.
- $((p \vee q) \vee r) \iff (p \vee (q \vee r))$ – associativity of $\vee$.
- $((p \wedge q) \wedge r) \iff (p \wedge (q \wedge r))$ – associativity of $\wedge$.
- $(\neg(p \vee q)) \iff (\neg p \wedge \neg q)$ – De Morgan's law ($\neg$ over $\vee$).
- $(\neg(p \wedge q)) \iff (\neg p \vee \neg q)$ – De Morgan's law ($\neg$ over $\wedge$).
- $(p \vee (q \wedge r)) \iff ((p \vee q) \wedge (p \vee r))$ – distributivity of $\vee$ over $\wedge$.
- $(p \wedge (q \vee r)) \iff ((p \wedge q) \vee (p \wedge r))$ – distributivity of $\wedge$ over $\vee$.
- $(p \implies q) \iff (q \impliedby p)$.
- $(p \iff q) \iff ((p \implies q) \wedge (p \impliedby q))$.

Quantifiers We now introduce constructs that are not part of propositional logic, but a higher-order logic called *predicate* logic. However, as they are useful for this course in intuiting properties in various contexts, we introduce and discuss them here. The constructs are called *quantifiers*, and they are useful when we want to make assertions that have variables in them.

An example of the use of a quantifier is the following: "every star is hot." Another way of saying the same thing, while explicating the use of a variable and a quantifier is: "for every star x, x is hot." The "for every" part is a quantifier, specifically the *universal* quantifier. The other quantifier of interest to use is the *existential* quantifier. An example of its use is: "there exists x such that x is a bird and x can fly." (More simply, in English we would say, "there exists a bird that can fly," or "some birds can fly.")

The notation we use for the universal quantifier is "$\forall$" and for the existential quantifier is "$\exists$." For example, we might write: "$\exists$ rational y such that $y^2 = 2$." As another example, "$\forall$ integer x, x^3 is an integer ." We can use the logical connectives $\neg, \vee$ and $\wedge$ along with quantifiers. For example, to express that there exists no rational y such that $y^2 = 2$, we could write: "$\neg(\exists$ rational y such that $y^2 = 2)$,"

In the context of that last example, it is useful to be able to intuit equivalent assertions. We could equivalently assert: "$\forall$ rational $y, \neg(y^2 = 2)$," for that example, or, "$\forall$ rational $y, y^2 \neq 2$," if we define the symbol "$\neq$" as the complement of "$=$." Indeed, following are the rules, in general, of negating an assertion with a quantifier. In the following, we assume that $p(x)$ is an assertion that involves the variable x.

- $\neg(\exists x, p(x)) \iff \forall x, \neg p(x)$.
- $\neg(\forall x, p(x)) \iff \exists x, \neg p(x)$.

We can quantify over more than one variable. For example: "$\forall$ positive integer a, $\exists$ real b such that $b = \sqrt{a}$." Note that, when different quantifiers are used, as in this example, their order matters: in general, "$\forall$ person a, $\exists$ person b such that b is a's mother" is not equivalent to "$\exists$ person b, such that, $\forall$ person a, b is a's mother"; the first formula asserts that every person has a mother, the second that there is a person who is mother to everyone (even herself).

Sometimes, when we use the same quantifier over multiple variables, we write one instance of a quantifier only, and not several. For example:

$$\forall \text{ real } a, b, (a \leq b \vee b \leq a)$$

When we really should write "$\forall$ real a, $\forall$ real $b \ldots$."

We have already been using quantifiers implicitly. For example, consider Claim 3 above. When we refer to p, q and r in the statement of the claim, what we really mean to say is, "for all propositions p, q and r, it is true that…" The "for all" quantifiers on each of p, q and r were left implicit in the statement of the claim.

Proof techniques

We now discuss proof techniques that are useful in this course, and in future, to you in your engineering profession. The mindset and systematic thinking that working out a proof develops is critical to one's success as an engineer. The kinds of proofs we develop, and the underlying mindsets and techniques we use, are not only of esoteric or theoretical interest. They have immediate, practical consequence. Also, the precise communication that such proofs require also are very valuable for one to develop as an engineer. Precise technical communication is an invaluable skill, that is highy prized not only in academia, but also industry and business settings. We return to this somewhat philosophical discussion once we have discussed the proof techniques we seek to impart as part of this course.

Logical deduction The overarching technique we use is logical deduction: going from a set of known or assumed statements to new statements, that are typically derived by logic implication. We have already seen some examples of this in our discussions on logic in this chapter.

Consider the following joke. Three logicians walk into a bar. The bartender asks, "would y'all like something to drink?" Logician 1 says, "I don't know." Logician 2 says, "I don't know." Logician 3 says, "yes."

The joke is a play on the wording of the bartender's question, specifically, her use of "all." She seems to be asking whether all three of the logicians want a drink. Presumably, each of Logicians 1 and 2 would like a drink. But they do not know yet as to whether all of them want a drink. Therefore, they are compelled to say, "I don't know." Logician 3 infers that the other two would each like a drink; otherwise, one of them would have said, "no." She knows that she wants a drink herself, and therefore says, "yes."

Imagine that Logician 3 had said, "no." Then, presumably Logicians 1 and 2 want a drink each, but Logician 3 does not. While this is admittedly a joke, it exercises logical deduction in a good way. Such logical deduction is at the foundations of every proof we carry out. Following are some specific strategies one could adopt to carry out a proof. Each strategy provides a kind of framework within which logical deduction is used. More than one strategy may be useful in carrying out a proof, and a proof does not require

any particular strategy to be adopted to be carried out successfully. It is important also to recognize when one has successfully carried out a proof; the strategy helps with this aspect as well.

Some of the strategies that arise in this course, and in future courses are:

- Case analysis: we enumerate, exhaustively, all possible cases that can occur, and prove each, in turn. Following is an example.

  Claim 4. *For any three natural numbers x, y, z, where $x + y = z$, if any two of x, y, z are divisible by 3, then so is the third.*

  Proof. By case analysis.

  1. x, y are divisible by 3. Then, $x = 3a, y = 3b$ for some natural numbers a, b. Then, because $z = x+y, z = 3(a+b)$, which implies that z is divisible by 3.
  2. x, z are divisible by 3. Then, $x = 3a, z = 3b$ for some natural numbers a, b. As y is a natural number, i.e., $y \geq 0$ and $x + y = z$, $b \geq a$. And, $y = 3(b - a)$. As $b \geq a$, $b - a$ is a natural number, and therefore y is a natural number that is divisible by 3.
  3. y, z are divisible by 3. This is identical to the previous case as x and y are interchangeable.

  □

  An interesting observation about the above claim is that its converse is not necessarily true. That is, for three natural numbers x, y, z with $x+y = z$, if one of them is divisible 3, it does not necessarily imply that the other two are as well. A *counterexample* can be used to establish this. A counterexample is $x = 1, y = 2, z = 3$.

- Contradiction: we recall the truth table for an implication, and observe that the only case such a proposition is false is when ϕ is true, and ψ is false. For a proof by contradiction of a proposition $\phi \implies \psi$, we assume that the premise, ϕ is true, and yet, the implication, ψ, is false. We then establish by logical deduction that something that is false must be true, or that something that is true must be false – this is the contradiction we deduce.

For example, consider the following claim, and its proof by contradiction.

Claim 5. *$\sqrt{2}$ is not rational.*

Proof. To perceive the statement the claim as an implication, we can rephrase it as: $x = \sqrt{2} \implies x$ is not a rational number.

For the purpose of contradiction, assume that $x = \sqrt{2}$, and x is rational. Then, $x = p/q$, where p and q are integers. We assume, without loss of generality, that p and q have only 1 as a common factor, i.e., p/q is in its simplest form. Then, $x^2 = 2 = p^2/q^2 \implies p^2 = 2q^2$.

Thus, p^2 is even. This implies that p is even, because if p is odd, then p is of the form $2x+1$ where x is an integer, and $(2x+1)^2 = 4x^2+4x+1$, which is odd. Thus, $p = 2y$, for some integer y.

Therefore, $p^2/2 = (2y)^2/2 = 2y^2 = q^2$. Thus, q^2 is even as well, and therefore q is even. Thus, both p and q are even, which means p/q is not in its simplest form, which is our desired contradiction.

□

Another example, which was on the final exam of the Spring'18 offering of the course is the following claim. We define an even number as follows: x is an even number if $x = 2y$, where y is an integer.

Claim 6. *If a, b, c are positive integers, then at least one of $a - b, b - c, c - a$ is even.*

An example is $a = 13, b = 8, c = 5$. Then, $c - a = -8$, which is even.

Proof. Assume, for the purpose of contradiction, that none of $a-b, b-c, c-a$ is even. Then, $a-b = 2k+1$ for some integer k, and $b-c = 2l+1$ for some integer l. then, $c-a = -(b-c+a-b) = -(2l+1+2k+1) = 2(-l-k+1)$, which is an integer because l, k are integers, and is even. This contradicts our assumption that $c - a$ is odd. □

- Contrapositive: recall that $(\phi \implies \psi) \iff (\neg\psi \implies \neg\phi)$; the two implications are contrapositives of one another. Given a claim $\phi \implies \psi$ a proof of the contrapositive proves, instead, $\neg\psi \implies \neg\phi$.

  Following is an example of proof by contrapositive.

Claim 7. *For x, y positive integers,* $\left(\sum_{i=1}^{x} i = \sum_{i=1}^{y} i\right) \implies (x = y)$.

Proof. We prove the contrapositive, that is, for x, y positive integers, $(x \neq y) \implies \left(\sum_{i=1}^{x} i \neq \sum_{i=1}^{y} i\right)$.

Given that $x \neq y$, either (i) $x > y$ or (ii) $x < y$. In case (i), $\left(\sum_{i=1}^{x} i\right) = \left(\sum_{i=1}^{y} i + \sum_{i=y+1}^{x} i\right) \geq \left(y + 1 + \sum_{i=1}^{y} i\right) > \left(1 + \sum_{i=1}^{y} i\right)$, because $x \geq y+1$ and $y > 0$. This implies that $\sum_{i=1}^{x} i \neq \sum_{i=1}^{y} i$, as desired.

Case (ii) is proven identically, by interchanging x and y. □

- Construction: this is typically for statements of the form "there exists..." That is, a natural way to prove that something exists is to construct, or present, one. For example, if we all agree on what an elephant is, and I am challenged to prove that elephants exist, I can simply produce and present an elephant. Following is an example.

  Claim 8. *Given any two real numbers, x, y such that $x < y$, there exists a real number z such that $x < z < y$.*

  Proof. By construction. Let $z = (x + y)/2$. Then z is real because the sum of two real numbers is real, and dividing a real by another that is not zero yields a real. To establish that $x < z < y$, we observe:

$$\begin{aligned} x < z < y &\impliedby x < \frac{x+y}{2} < y \\ &\impliedby 2x < x + y < 2y \\ &\impliedby x + x < x + y < y + y \\ &\impliedby x < y \end{aligned}$$

  □

The above proof demonstrates a useful strategy: to begin with what we seek to prove, and then work backwards to a sufficient condition for that to be true, in this case, $x < y$, which we know to be true.

- Induction: a proof by induction is usually put to use when we have a statement that involves a universal quantifier, for a sequence of items, for example, all natural numbers. A proof by induction is structured as follows:
  - We first prove that the statement is true for the *base case*. The base case is the statement for the first natural number, 0.
  - We then prove the *step*, i.e., the following implication: if the statement is true for all natural numbers, $0, 1, \ldots, i-1$, then the statement is true for the natural number i.

  Together, the two steps above prove the statement for all items in the sequence, for example, every natural number. This is because proving the (i) base case, i.e., the statement for 0, and, (ii) the step, implies that the statement is true for the second natural number, 1. This, with the step, in turn implies that the statement is true for 2. And, 3, and so on, for all natural numbers. Following is an example.

Claim 9. $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$.

Proof. By induction on n.

Base case: $n = 1$. When $n = 1$, the left hand side is 1. And the right hand-side is $\frac{1\times 2}{2} = 1$. Thus, we have proved that the statement is true for the base case.

Step: we adopt the induction assumption, that the statement is true for all $n = 1, 2, \ldots, i-1$, for some $i \geq 2$. Under that premise, we seek to prove the statement for $n = i$. We observe:

$$\begin{aligned} 1 + 2 + \ldots + i - 1 + i &= \frac{(i-1)i}{2} + i \qquad \because \text{induction assumption} \\ &= \frac{i^2 - i + 2i}{2} \\ &= \frac{i^2 + i}{2} = \frac{i(i+1)}{2} \end{aligned}$$

Thus, we have proven the base case and the step, and therefore we have successfully carried out our proof by induction on n.

□

As the base case, we have proved that the statement is true when $n = 1$. As a consequence of proving the step, then, we have proved that the statement is true for $n = 2$. And with that, and as a consequence of the step, we have proved that the statement is true for $n = 3$. And so on.

We now carry out several proofs as examples to demonstrate the above strategies. We begin with a problem from the final exam of the Spring'18 offering of the course.

Claim 10. *For every non-negative integer $n \geq 12$, there exist non-negative integers m_1, m_2 such that $n = 4m_1 + 5m_2$.*

Proof. By induction on n.

Base cases: we prove the statement for the following cases: $n = 12, 13, 14, 15$. The reason we consider several base cases becomes apparent once we get in to proving the step. We observe:

- $12 = 4 \times 3 + 5 \times 0$.
- $13 = 4 \times 2 + 5 \times 1$.
- $14 = 4 \times 1 + 5 \times 2$.
- $15 = 4 \times 0 + 5 \times 3$.

Step: we assume that the assertion is true for all $n = 12, 13, \ldots, i - 1$ for some $i \geq 13$. For $n = i$, we first observe that $i = i - 1 + 1 = 4k_1 + 5k_2 + 1$, for some non-negative integers k_1, k_2, from the induction assumption. We do a case analysis.

Case (i): $k_1 > 0$. Then, $i = 4k_1 + 5k_2 + 1 = 4(k_1 - 1) + 5(k_2 + 1)$.

Case (ii): $k_1 = 0$. Then, because $i > 12$, $k_2 \geq 3$. Then, $i = 5k_2 + 1 = 4 \times 4 + 5(k_2 - 3)$.

The reason we prove several base cases is to address Case (ii) of the step. Because the smallest n for which $k_2 \geq 3$ is $n = 15$. By addressing several base cases, we ensure that our proof is indeed correct, i.e., that we can indeed make the inductive argument. □

Claim 11. *For every non-negative integer n, exactly one of the following is true:*

- *there exists a non-negative integer m such that $n = 3m$.*
- *there exists a non-negative integer m such that $n = 3m + 1$.*
- *there exists a non-negative integer m such that $n = 3m + 2$.*

We need to be careful here in that the statement says that exactly one of those cases is true. That is, for a particular n, one of the cases is true, and neither of the others is true. We need to prove both those properties.

Proof. By induction on n. Again, we are careful to address several base cases.

Base cases: for each of $n = 0, 1, 2$, we prove the first part by construction, i.e., by producing an m that demonstrates that the statement is true.

For $n = 0$, we observe that $0 = 3 \times 0$, i.e., $m = 0$, which proves that the statement is true. For $n = 1$, we again propose $m = 0$, and observe that $n = 1 = 3 \times 0 + 1$. And for $n = 2$, we propose $m = 0$, and observe that $n = 2 = 3 \times 0 + 2$. Thus, we have shown one part of the statement for each of $n = 0, 1, 2$, which is that there exists such an m.

We now prove the other part of the statement: that given that $0 = 3m$ for some m, then it can be neither $3m' + 1$ nor $3m' + 2$ for any non-negative integer m'. Suppose, for the purpose of contradiction, there exists such an m', that is, $0 = 3m' + 1$. Then, $m' = -1/3$, which contradicts the assumption that m' is a non-negative integer. Similarly, $0 = 3m' + 2 \implies m' = -2/3$, again a contradiction.

And similarly, if $1 = 3m'$, then $m' = 1/3$ and if $1 = 3m' + 2$, then $m' = -1/3$, in each case a contradiction to the assumption that m' is a non-negative integer. And finally, if $2 = 3m'$, then $m' = 2/3$, and if $2 = 3m' + 1$, then $m' = 1/3$.

Step: we assume that the statement is true for all $n = 0, 1, 2, \ldots, i-1$ for some $i \geq 1$. For $n = i$, we do a case analysis, and in each case, produce an m.

- if $i - 1 = 3m$ for some non-negative integer m, then, $i = 3m + 1$.
- if $i - 1 = 3m + 1$ for some non-negative integer m, then, $i = 3m + 2$.
- if $i - 1 = 3m + 2$ for some non-negative integer m, then, $i = 3(m + 1)$. And because m is a non-negative integer, so is $m + 1$.

To establish that no other case applies, assume that a non-negative integer m' exists that corresponds to one of the other cases, for the purpose of contradiction. We again do a case analysis.

- if $i = 3m$ and $i = 3m' + 1$, then $m' = m - 1/3$, which is a contradiction to the assumption that m' is a non-negative integer. And if $i = 3m' + 2$, then $m' = m - 2/3$, which is a similar contradiction.
- if $i = 3m + 1$ and $i = 3m'$, then $m' = m + 1/3$, and if $i = 3m' + 2$, then $m' = m - 1/3$, each of which is a contradiction.
- if $i = 3m + 2$ and $i = 3m'$, then $m' = m + 2/3$, and if $i = 3m' + 1$, then $m' = m + 1/3$, both of which contradict our assumption that m' is a non-negative integer.

□

We now consider a proof by induction for a statement that is obviously not true. The statement is: all horses have the same colour. The proof is as follows. For the base case, pick a horse. Obviously it is the same colour as itself. Therefore, the base case has been proved. The induction assumption is that given up to $n = i - 1$ horses, for some $i \geq 2$, they all have the same colour. Now consider that we are given $n = i$ horses. We pick some horse, and temporarily remove it from the set. Then we are left with $i - 1$ horses which, by the induction assumption all have the same colour. We now temporarily remove one of those $i - 1$ horses from the set, and add back in

the horse that we first removed. Again, we are left with $i-1$ horses which, by the induction assumption must all have the same colour.

A flaw in the above proof is in the manner in which we prove the step. While it is certainly ok to remove a horse, call it H, from the set and then assert that the remainder all have the same colour, what we now need to do is prove that H has the same colour as the other $i-1$ horses. We cannot again appeal to the induction assumption to do that, as the above flawed proof does.

We now present one more correct example of proof by induction. In the following claim, we address a situation that there appears to be more than one choice for the parameter on which we carry out induction.

Claim 12. *Suppose n is a natural number whose digits, in order of most-to least-significant, are $n_{k-1}\, n_{k-2}\, \ldots, n_0$, where each n_i is one of $0, \ldots, 9$. If the sum of the digits of n, $S_n = \sum_{i=0}^{k-1} n_i$, is divisible by 3, then n is divisible by 3.*

An example is $n = 809173$. Then, $S_n = 28$, which is not divisible by 3. Therefore, from the statement in the claim, we cannot infer anything as to whether n is divisible by 3. On the other hand, the digits of 82907370 add up to 36, and therefore, if the claim is true, then 82907370 is divisible by 3.

We emphasize that the implication in the statement goes in one direction only. "...if S_n is divisible by 3, then n is divisible by 3..." It says nothing about what S_n may be if n is divisible by 3.

The above claim presents an example of where if we choose to carry out a proof by induction, then we need to clearly say on what parameter we carry out induction. For the above claim, there appear to be at least two choices: induction on n, and induction on k. In the following proof, we carry out induction on k, i.e., the number of digits when we write n in decimal.

Proof. Base case: $k = 1$. Then, $n = n_0 = S_n$, i.e., n has only one digit. Then, for S_n to be divisible by 3, S_n must be one of $3, 6$ or 9. In each case, because $n = S_n$, we observe that n is divible by 3 as well.

Step: our induction assumption is that given any n that has $k = 1, \ldots, i-1$ digits, for some $i \geq 2$, if S_n is divisible by 3, then so is n. We need to

now prove that given some n of i digits, if S_n is divisible by 3, then so is n. Henceforth, we use the notation $(\,)_{10}$ to indicate when we write a number in base-10, i.e., its digits from most- to least-significant.

We have $n = (n_{i-1}\, n_{i-2} \ldots n_0)_{10}$. Therefore, $n = 10^{i-1}n_{i-1} + 10^{i-2}n_{i-2} + \ldots + 10^0 n_0 = 10^{i-1}n_{i-1} + (n_{i-2} \ldots n_0)_{10}$. Also, $S_n = \sum_{j=0}^{i-1} n_j = n_{i-1} + \sum_{j=0}^{i-2} n_j$. We do a case analysis on $\sum_{j=0}^{i-2} n_j$ as to whether it is divisible by 3. We appeal often to Claim 4. Recall that that claim is: given three natural numbers x, y, z such that $x + y = z$ and any two are divisible by 3, then so is the third.

- Suppose $\sum_{j=0}^{i-2} n_j$ is divisible by 3. Then, for S_n to be divisible by 3, n_{i-1} must be divisible by 3 by Claim 4. That is, $n_{i-1} = 3a$ for some natural number a. Then, $n = 10^{i-1} \times 3a + (n_{i-2} \ldots n_0)_{10}$. As $\sum_{j=0}^{i-2} n_j$ is divisible by 3, by the induction assumption, $(n_{i-2} \ldots n_0)_{10}$ is divisible by 3. Therefore, by Claim 4, $n = 10^{i-1} \times 3a + (n_{i-2} \ldots n_0)_{10}$ is divisible by 3, because it is the sum of two numbers, each of which is divisible by 3.

- Suppose $\sum_{j=0}^{i-2} n_j$ is not divisible by 3. Then, $\sum_{j=0}^{i-2} n_j = 3a + b$, for some natural number a, and for b either 1 or 2. We now do a case analysis of those two cases for b.

  - If $b = 1$, then $n_{i-1} = 3a' + 2$ for some natural number a', because otherwise, S_n is not divisible by 3. And we have:

$$\begin{aligned} n &= 10^{i-1}n_{i-1} + (n_{i-2} \ldots n_0)_{10} \\ &= 10^{i-1}(3a' + 2) + (n_{i-2} \ldots n_0)_{10} \\ &= 10^{i-1} \times 3a' + 10^{i-2} \times 20 + (n_{i-2} \ldots n_0)_{10} \end{aligned}$$

    Now, we do a further case analysis on n_{i-2}:

* If $n_{i-2} = 0$, then, we choose to write n as:

$$n = 10^{i-1} \times 3a' + 10^{i-2} \times 18 + (2\, n_{i-3} \ldots n_0)_{10}$$

Now, each of $10^{i-1} \times 3a'$ and $10^{i-2} \times 18$ is divisible by 3. And the digits of $(2\, n_{i-3} \ldots n_0)_{10}$ are divisible by 3, because $\sum_{j=0}^{i-2} n_j = 3a + 1$. Therefore, by the induction assumption, $(2\, n_{i-3} \ldots n_0)$ is divisible by 3. Thus, n is the sum of three numbers, each of which is divisible by 3, and therefore n is divisible by 3.

* If $n_{i-2} > 0$, then, we choose to write n as:

$$n = 10^{i-1} \times 3a' + 10^{i-2} \times 21 + ((n_{i-2} - 1)\, n_{i-3} \ldots n_0)_{10}$$

Again, n is the sum of three numbers each of which is divisible by 3.

– If $b = 2$, then $n_{i-1} = 3a' + 1$ for some natural number a', because otherwise, S_n is not divisible by 3. And we have:

$$\begin{aligned} n &= 10^{i-1} n_{i-1} + (n_{i-2} \ldots n_0)_{10} \\ &= 10^{i-1}(3a' + 1) + (n_{i-2} \ldots n_0)_{10} \\ &= 10^{i-1} \times 3a' + 10^{i-2} \times 10 + (n_{i-2} \ldots n_0)_{10} \end{aligned}$$

As before, we do a further case analysis on n_{i-2}:

* If $n_{i-2} = 0$ or $n_{i-2} = 1$, then we choose to write n as:

$$n = 10^{i-1} \times 3a' + 10^{i-2} \times 9 + ((n_{i-2} + 1)\, n_{i-3} \ldots n_0)_{10}$$

And n is the sum of three numbers each of which is divisible by 3.

* If $n_{i-2} \geq 2$, then we choose to write n as:

$$n = 10^{i-1} \times 3a' + 10^{i-2} \times 12 + ((n_{i-2} - 2)\, n_{i-3} \ldots n_0)_{10}$$

And n is the sum of three numbers each of which is divisible by 3.

□

Disproof

Sometimes we are faced with a statement that we don't know to be true or to be false. We may need to *find out* whether it's true or false. In such a case, we can see whether we can prove it, or disprove it.

Consider the following example.

Claim 13. *For every natural number n, $n^2 - n + 11$ is prime.*

If we simply try to prove this statement, we will never succeed. But we may succeed in disproving it: it turns out that the above claim is false. That is, there exists natural n such that $n^2 - n + 11$ is not prime.

Such an n is 11, and this is called a *counterexample* to the claim: an example of a specific n for which the claim does not hold. Producing a counterexample is an effective way of refuting a statement of the form "for all ..."

For another disproof by counterexample, consider the statement, "no mammal lays eggs." This can be seen as the negation of a statement with the "there exists" quantifier. Which can in turn be rephrased as a statement with "for all..."

$$\begin{aligned} \text{No mammal lays eggs} &\iff \nexists \text{ a mammal that lays eggs} \\ &\iff \neg(\exists \text{ a mammal that lays eggs}) \\ &\iff \forall \text{ mammals } m, m \text{ does not lay eggs} \end{aligned}$$

As a counterexample to the latter statement, we could present the platypus, which is an egg-laying mammal. A note of caution: sometimes it is not obvious that something that is presented as a counterexample is indeed a counterexample. In such a situation, we need to prove that it is indeed a counterexample. For example, we need to prove that the platypus that we present as a counterexample is indeed a mammal, and does lay eggs.

For our counterexample for Claim 13, as proof that $n = 11$ is indeed a valid counterexample, we would observe that $11^2 - 11 + 11 = 121$, which is not prime because it has a divisor, 11, which is neither itself nor 1. The proof of Claim 5 of Chapter 2 establishes the non-obvious fact that the square root of 2 is a valid counterexample to the claim that all real numbers are rational.

Why is the presentation of a counterexample a valid way of disproving a statement of the form "for all ..."? The reason is that we are proving the negation of the statement. That is, to prove that a statement S is false, we prove $\neg S$ to be true. Thus, if $P(x)$ is a statement about x, and $S = \forall x, P(x)$, then:

$$\neg S \iff \neg(\forall x, P(x)) \iff \exists x, \neg P(x)$$

And then, a counterexample is a proof by construction of $\exists x, \neg P(x)$.

We consider one more example of a claim that we are able to disprove by counterexample.

Claim 14. *For all sets A, B and C, $A \times C = B \times C \implies A = B$.*

As a counterexample, pick $C = \emptyset$ and A, B be any sets such that $A \neq B$.

To disprove a statement of a form other than "for all ..." we simply negate the statement, and prove that this negation is true. For example, to disprove a statement of the form "there exists ...", we need to prove a statement of the form "for all ..." That is:

Let $S = \exists x, P(x)$
Then, $\neg S = \neg(\exists x, P(x)) \iff \forall x, \neg P(x)$

Claim 15. *There exist primes p, q such that $p - q = 513$.*

The above claim is false. Its negation is:

For all primes p, q, $p - q \neq 513$

We can prove this by contradiction. Suppose there exist primes p, q such that $p - q = 513$. (Observe that this is exactly the statement of Claim 15.) Then, one of p, q is even and the other is odd. We now do a case-analysis. (i) Suppose p is even, and q is odd. Then, $p = 2$ as that is the only even prime. Then, $q = -511$, which contradicts the assumption that q is prime. (We adopt the customary condition that for a number to be prime, it must

be a natural number, i.e., ≥ 1.) (ii) Suppose q is even and p is odd. Then, $q = 2$ and $p = 515$, which contradicts the assumption that p is prime.

We now consider a more complex example, a statement that involves two quantifiers. This example illustrates the utility of first carefully negating the statement, and then choosing a strategy when trying to disprove the original statement.

Claim 16. $\forall m \in \mathbb{Z}, \exists n \in \mathbb{N}, \left|\frac{1}{m} - \frac{1}{n}\right| > \frac{1}{2}$.

The above statement is not true. Its negation, which we seek to prove, is:

$$\exists m \in \mathbb{Z}, \forall n \in \mathbb{N}, \left|\frac{1}{m} - \frac{1}{n}\right| \leq \frac{1}{2}$$

We can prove this statement by construction of a suitable m, and then proving that for that choice of m, the "$\forall n \ldots$" part is true. Choose $m = 2$. Then, we perform a case-analysis on n.

When $n = 1$, $\left|\frac{1}{2} - 1\right| = \frac{1}{2} \leq \frac{1}{2}$.

When $n = 2$, $\left|\frac{1}{2} - \frac{1}{2}\right| = 0 \leq \frac{1}{2}$.

When $n \geq 3$, $0 < \frac{1}{n} < \frac{1}{2}$. Therefore, $\left|\frac{1}{2} - \frac{1}{n}\right| = \frac{1}{2} - \frac{1}{n} < \frac{1}{2}$.

We conclude with an example of a logical implication which does not hold. We use this example to illustrate the fact that when a statement that is an implication is false, this means that there is some assignment of truth-values to the constituent propositions that causes the implication not to hold. For some other assignments, it may or may not hold. And of course, for a proposition to be true, it must be true for all truth-assignments of its constituent propositions.

Claim 17. $(p \implies q) \implies ((p \vee r) \implies q)$

To disprove the above claim, we observe that when p is false, q is false and r is true, the statement is not true. We observe that for some other truth-assignments, the statement is true; for example, $p = \mathsf{true}, q = \mathsf{false}, r = \mathsf{true}$. But that is immaterial to the fact that the claim is false.

Another example of the use of the kind of logic and proof techniques is the following claim and proof. In the proof, we directly use " $\iff$." We could, instead, first establish $\overline{(\overline{A})} \subseteq A$, and then $\overline{(\overline{A})} \supseteq A$. The symbol "\" denotes set difference.

Claim 18. $\overline{(\overline{A})} = A$.

Proof.

$$\begin{aligned}
x \in \overline{(\overline{A})} &\iff x \in \mathcal{U} \setminus \overline{A} \\
&\iff x \in \mathcal{U} \wedge x \notin \overline{A} \\
&\iff x \in \mathcal{U} \wedge x \notin (\mathcal{U} \setminus A) \\
&\iff x \in \mathcal{U} \wedge \neg(x \in \mathcal{U} \setminus A) \\
&\iff x \in \mathcal{U} \wedge \neg(x \in \mathcal{U} \wedge \neg(x \in A)) \\
&\iff x \in \mathcal{U} \wedge (x \notin \mathcal{U} \vee x \in A) \\
&\iff (x \in \mathcal{U} \wedge x \notin \mathcal{U}) \vee (x \in \mathcal{U} \wedge x \in A) \\
&\iff \mathsf{false} \vee (x \in \mathcal{U} \wedge x \in A) \\
&\iff x \in \mathcal{U} \wedge x \in A \\
&\iff x \in \mathcal{U} \cap A \iff x \in A
\end{aligned}$$

□

As another somewhat more general example of the use of logic in the context of sets, we discuss Russel's paradox, which points out that the set builder notation should be used with care.

Russell's paradox A set is a collection of items. A set itself can be perceived as an item. Therefore, it is possible to specify a set of sets. For example, $\{\{1\}, \emptyset, \{1, 2, 3, 4, 5\}\}$ is a set of sets of integers, which has three members. An immediate question that then arises is: can a set be a member of itself? It does not seem meaningful to allow this, and therefore we may mandate that no set is allowed to be a member of itself.

However, it turns out that this by itself does not preclude contradictions that can occur in the specification of a set. A particular contradiction is Russell's

paradox, which is demonstrated by the following specification of a set using set-builder notation.

$$\text{Let } S = \{x \mid x \text{ is a set with the property that } x \notin x\}$$

That is, S is the set of all sets that do not contain themselves. Now, we ask: does S contain itself?

- If the answer is 'yes,' then:

  $$S \in S \implies S \text{ is a set that does not contain itself} \implies S \notin S$$

  Thus, we have a contradiction.

- If the answer is 'no,' then:

  $$S \notin S \implies S \text{ is a set that does not contain itself} \implies S \in S$$

  Thus, we again have a contradiction.

A thorough discussion on "clean" specifications of sets and other constructs is beyond the scope of this course. The above discussion on Russell's paradox reveals, however, that care must be taken. A quick "hack" is to restrict the manner in which the set-builder notation is used. We require that when specifying a set using the set-builder notation, it must look like the following:

$$\{x \in A \mid \text{conditions on } x\}$$

That is, we must specify of what superset A this set being specified is a subset. And the conditions that appear after "$\mid$" are then used to specify which members of A are members of this set. Under these requirements, the earlier specification, $S = \{x \mid x \notin x\}$ is no longer allowed.

And if we specify, for example, $S = \{x \in A \mid x \notin x\}$, we no longer have a paradox. Because suppose $S = \{x \in A \mid x \notin x\}$ is our specification of S, and we again ask: is $S \in S$?

- If the answer is 'yes,' then:

$$S \in S \implies S \in A \wedge S \notin S \implies S \notin S$$

  Thus, we have a contradiction.

- If the answer is 'no,' then:

$$S \notin S \implies S \notin A \vee (S \in A \wedge S \notin S)$$

  Now, if $S \in A$, then $S \in A \wedge S \notin S \implies S \in S$, a contradiction.

Thus, we have a possibility without a contradiction, and that is that $S \notin A$. Which implies $S \notin S$, and the answer to the question "is $S \in S$?" is "no."

Discrete probability

From Cormen, et al., "Introduction to Algorithms."

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

C.1 Counting

Counting theory tries to answer the question "How many?" without actually enumerating how many. For example, we might ask, "How many different n-bit numbers are there?" or "How many orderings of n distinct elements are there?" In this section, we review the elements of counting theory. Since some of the material assumes a basic understanding of sets, the reader is advised to start by reviewing the material in Section B.1.

Rules of sum and product

A set of items that we wish to count can sometimes be expressed as a union of disjoint sets or as a Cartesian product of sets.

The ***rule of sum*** says that the number of ways to choose an element from one of two *disjoint* sets is the sum of the cardinalities of the sets. That is, if A and B are two finite sets with no members in common, then $|A \cup B| = |A| + |B|$, which

follows from equation (B.3). For example, each position on a car's license plate is a letter or a digit. The number of possibilities for each position is therefore $26 + 10 = 36$, since there are 26 choices if it is a letter and 10 choices if it is a digit.

The ***rule of product*** says that the number of ways to choose an ordered pair is the number of ways to choose the first element times the number of ways to choose the second element. That is, if A and B are two finite sets, then $|A \times B| = |A| \cdot |B|$, which is simply equation (B.4). For example, if an ice-cream parlor offers 28 flavors of ice cream and 4 toppings, the number of possible sundaes with one scoop of ice cream and one topping is $28 \cdot 4 = 112$.

Strings

A ***string*** over a finite set S is a sequence of elements of S. For example, there are 8 binary strings of length 3:

$$000, 001, 010, 011, 100, 101, 110, 111 .$$

We sometimes call a string of length k a ***k-string***. A ***substring*** s' of a string s is an ordered sequence of consecutive elements of s. A ***k-substring*** of a string is a substring of length k. For example, 010 is a 3-substring of 01101001 (the 3-substring that begins in position 4), but 111 is not a substring of 01101001.

A k-string over a set S can be viewed as an element of the Cartesian product S^k of k-tuples; thus, there are $|S|^k$ strings of length k. For example, the number of binary k-strings is 2^k. Intuitively, to construct a k-string over an n-set, we have n ways to pick the first element; for each of these choices, we have n ways to pick the second element; and so forth k times. This construction leads to the k-fold product $n \cdot n \cdots n = n^k$ as the number of k-strings.

Permutations

A ***permutation*** of a finite set S is an ordered sequence of all the elements of S, with each element appearing exactly once. For example, if $S = \{a, b, c\}$, there are 6 permutations of S:

$$abc, acb, bac, bca, cab, cba .$$

There are $n!$ permutations of a set of n elements, since the first element of the sequence can be chosen in n ways, the second in $n - 1$ ways, the third in $n - 2$ ways, and so on.

A ***k-permutation*** of S is an ordered sequence of k elements of S, with no element appearing more than once in the sequence. (Thus, an ordinary permutation is just an n-permutation of an n-set.) The twelve 2-permutations of the set $\{a, b, c, d\}$ are

$$ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc .$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

The number of k-permutations of an n-set is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}\,, \tag{C.1}$$

since there are n ways of choosing the first element, $n-1$ ways of choosing the second element, and so on until k elements are selected, the last being a selection from $n-k+1$ elements.

Combinations

A ***k-combination*** of an n-set S is simply a k-subset of S. For example, there are six 2-combinations of the 4-set $\{a, b, c, d\}$:

$$ab, ac, ad, bc, bd, cd\ .$$

(Here we use the shorthand of denoting the 2-set $\{a, b\}$ by ab, and so on.) We can construct a k-combination of an n-set by choosing k distinct (different) elements from the n-set.

The number of k-combinations of an n-set can be expressed in terms of the number of k-permutations of an n-set. For every k-combination, there are exactly $k!$ permutations of its elements, each of which is a distinct k-permutation of the n-set. Thus, the number of k-combinations of an n-set is the number of k-permutations divided by $k!$; from equation (C.1), this quantity is

$$\frac{n!}{k!\,(n-k)!}\,. \tag{C.2}$$

For $k = 0$, this formula tells us that the number of ways to choose 0 elements from an n-set is 1 (not 0), since $0! = 1$.

Binomial coefficients

We use the notation $\binom{n}{k}$ (read "n choose k") to denote the number of k-combinations of an n-set. From equation (C.2), we have

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

This formula is symmetric in k and $n-k$:

$$\binom{n}{k} = \binom{n}{n-k}\,. \tag{C.3}$$

These numbers are also known as ***binomial coefficients***, due to their appearance in the ***binomial expansion***:

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} . \tag{C.4}$$

A special case of the binomial expansion occurs when $x = y = 1$:

$$2^n = \sum_{k=0}^{n} \binom{n}{k} .$$

This formula corresponds to counting the 2^n binary n-strings by the number of 1's they contain: there are $\binom{n}{k}$ binary n-strings containing exactly k 1's, since there are $\binom{n}{k}$ ways to choose k out of the n positions in which to place the 1's.

There are many identities involving binomial coefficients. The exercises at the end of this section give you the opportunity to prove a few.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

C.2 Probability

Probability is an essential tool for the design and analysis of probabilistic and randomized algorithms. This section reviews basic probability theory.

We define probability in terms of a ***sample space*** S, which is a set whose elements are called ***elementary events***. Each elementary event can be viewed as a possible outcome of an experiment. For the experiment of flipping two distinguishable coins, we can view the sample space as consisting of the set of all possible 2-strings over $\{\text{H}, \text{T}\}$:

$$S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\} \ .$$

An ***event*** is a subset[1] of the sample space S. For example, in the experiment of flipping two coins, the event of obtaining one head and one tail is $\{\text{HT}, \text{TH}\}$. The event S is called the ***certain event***, and the event $\emptyset$ is called the ***null event***. We say that two events A and B are ***mutually exclusive*** if $A \cap B = \emptyset$. We sometimes treat an elementary event $s \in S$ as the event $\{s\}$. By definition, all elementary events are mutually exclusive.

Axioms of probability

A ***probability distribution*** $\Pr\{\}$ on a sample space S is a mapping from events of S to real numbers such that the following ***probability axioms*** are satisfied:

1. $\Pr\{A\} \geq 0$ for any event A.
2. $\Pr\{S\} = 1$.

[1] For a general probability distribution, there may be some subsets of the sample space S that are not considered to be events. This situation usually arises when the sample space is uncountably infinite. The main requirement is that the set of events of a sample space be closed under the operations of taking the complement of an event, forming the union of a finite or countable number of events, and taking the intersection of a finite or countable number of events. Most of the probability distributions we shall see are over finite or countable sample spaces, and we shall generally consider all subsets of a sample space to be events. A notable exception is the continuous uniform probability distribution, which will be presented shortly.

3. $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}$ for any two mutually exclusive events A and B. More generally, for any (finite or countably infinite) sequence of events $A_1, A_2, \ldots$ that are pairwise mutually exclusive,

$$\Pr\left\{\bigcup_i A_i\right\} = \sum_i \Pr\{A_i\} \ .$$

We call $\Pr\{A\}$ the ***probability*** of the event A. We note here that axiom 2 is a normalization requirement: there is really nothing fundamental about choosing 1 as the probability of the certain event, except that it is natural and convenient.

Several results follow immediately from these axioms and basic set theory (see Section B.1). The null event $\emptyset$ has probability $\Pr\{\emptyset\} = 0$. If $A \subseteq B$, then $\Pr\{A\} \leq \Pr\{B\}$. Using $\overline{A}$ to denote the event $S - A$ (the ***complement*** of A), we have $\Pr\{\overline{A}\} = 1 - \Pr\{A\}$. For any two events A and B,

$$\begin{aligned} \Pr\{A \cup B\} &= \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} && \text{(C.12)} \\ &\leq \Pr\{A\} + \Pr\{B\} \ . && \text{(C.13)} \end{aligned}$$

In our coin-flipping example, suppose that each of the four elementary events has probability $1/4$. Then the probability of getting at least one head is

$$\begin{aligned} \Pr\{\text{HH}, \text{HT}, \text{TH}\} &= \Pr\{\text{HH}\} + \Pr\{\text{HT}\} + \Pr\{\text{TH}\} \\ &= 3/4 \ . \end{aligned}$$

Alternatively, since the probability of getting strictly less than one head is $\Pr\{\text{TT}\} = 1/4$, the probability of getting at least one head is $1 - 1/4 = 3/4$.

Discrete probability distributions

A probability distribution is ***discrete*** if it is defined over a finite or countably infinite sample space. Let S be the sample space. Then for any event A,

$$\Pr\{A\} = \sum_{s \in A} \Pr\{s\} \ ,$$

since elementary events, specifically those in A, are mutually exclusive. If S is finite and every elementary event $s \in S$ has probability

$$\Pr\{s\} = 1/|S| \ ,$$

then we have the ***uniform probability distribution*** on S. In such a case the experiment is often described as "picking an element of S at random."

As an example, consider the process of flipping a ***fair coin***, one for which the probability of obtaining a head is the same as the probability of obtaining a tail, that is, $1/2$. If we flip the coin n times, we have the uniform probability distribution

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

defined on the sample space $S = \{\text{H}, \text{T}\}^n$, a set of size 2^n. Each elementary event in S can be represented as a string of length n over $\{\text{H}, \text{T}\}$, and each occurs with probability $1/2^n$. The event

$A = \{\text{exactly } k \text{ heads and exactly } n - k \text{ tails occur}\}$

is a subset of S of size $|A| = \binom{n}{k}$, since there are $\binom{n}{k}$ strings of length n over $\{\text{H}, \text{T}\}$ that contain exactly k H's. The probability of event A is thus $\Pr\{A\} = \binom{n}{k}/2^n$.

Conditional probability and independence

Sometimes we have some prior partial knowledge about the outcome of an experiment. For example, suppose that a friend has flipped two fair coins and has told you that at least one of the coins showed a head. What is the probability that both coins are heads? The information given eliminates the possibility of two tails. The three remaining elementary events are equally likely, so we infer that each occurs with probability $1/3$. Since only one of these elementary events shows two heads, the answer to our question is $1/3$.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Conditional probability formalizes the notion of having prior partial knowledge of the outcome of an experiment. The ***conditional probability*** of an event A given that another event B occurs is defined to be

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \tag{C.14}$$

whenever $\Pr\{B\} \neq 0$. (We read "$\Pr\{A \mid B\}$" as "the probability of A given B.") Intuitively, since we are given that event B occurs, the event that A also occurs is $A \cap B$. That is, $A \cap B$ is the set of outcomes in which both A and B occur. Since the outcome is one of the elementary events in B, we normalize the probabilities of all the elementary events in B by dividing them by $\Pr\{B\}$, so that they sum to 1. The conditional probability of A given B is, therefore, the ratio of the probability of event $A \cap B$ to the probability of event B. In the example above, A is the event that both coins are heads, and B is the event that at least one coin is a head. Thus, $\Pr\{A \mid B\} = (1/4)/(3/4) = 1/3$.

Two events are ***independent*** if

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\} \ , \tag{C.15}$$

which is equivalent, if $\Pr\{B\} \neq 0$, to the condition

$$\Pr\{A \mid B\} = \Pr\{A\} \ .$$

For example, suppose that two fair coins are flipped and that the outcomes are independent. Then the probability of two heads is $(1/2)(1/2) = 1/4$. Now suppose that one event is that the first coin comes up heads and the other event is that the coins come up differently. Each of these events occurs with probability $1/2$, and the probability that both events occur is $1/4$; thus, according to the definition of independence, the events are independent—even though one might think that both events depend on the first coin. Finally, suppose that the coins are welded together so that they both fall heads or both fall tails and that the two possibilities are equally likely. Then the probability that each coin comes up heads is $1/2$, but the probability that they both come up heads is $1/2 \neq (1/2)(1/2)$. Consequently, the event that one comes up heads and the event that the other comes up heads are not independent.

A collection $A_1, A_2, \ldots, A_n$ of events is said to be ***pairwise independent*** if

$$\Pr\{A_i \cap A_j\} = \Pr\{A_i\}\Pr\{A_j\}$$

for all $1 \leq i < j \leq n$. We say that the events of the collection are ***(mutually) independent*** if every k-subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ of the collection, where $2 \leq k \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, satisfies

$$\Pr\{A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\} = \Pr\{A_{i_1}\}\Pr\{A_{i_2}\}\cdots\Pr\{A_{i_k}\} \ .$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

For example, suppose we flip two fair coins. Let A_1 be the event that the first coin is heads, let A_2 be the event that the second coin is heads, and let A_3 be the event that the two coins are different. We have

$$\begin{aligned}
\Pr\{A_1\} &= 1/2\,, \\
\Pr\{A_2\} &= 1/2\,, \\
\Pr\{A_3\} &= 1/2\,, \\
\Pr\{A_1 \cap A_2\} &= 1/4\,, \\
\Pr\{A_1 \cap A_3\} &= 1/4\,, \\
\Pr\{A_2 \cap A_3\} &= 1/4\,, \\
\Pr\{A_1 \cap A_2 \cap A_3\} &= 0\,.
\end{aligned}$$

Since for $1 \le i < j \le 3$, we have $\Pr\{A_i \cap A_j\} = \Pr\{A_i\}\Pr\{A_j\} = 1/4$, the events A_1, A_2, and A_3 are pairwise independent. The events are not mutually independent, however, because $\Pr\{A_1 \cap A_2 \cap A_3\} = 0$ and $\Pr\{A_1\}\Pr\{A_2\}\Pr\{A_3\} = 1/8 \neq 0$.

Bayes's theorem

From the definition of conditional probability (C.14) and the commutative law $A \cap B = B \cap A$, it follows that for two events A and B, each with nonzero probability,

$$\begin{aligned}
\Pr\{A \cap B\} &= \Pr\{B\}\Pr\{A \mid B\} \\
&= \Pr\{A\}\Pr\{B \mid A\}\,.
\end{aligned} \tag{C.16}$$

Solving for $\Pr\{A \mid B\}$, we obtain

$$\Pr\{A \mid B\} = \frac{\Pr\{A\}\Pr\{B \mid A\}}{\Pr\{B\}}\,, \tag{C.17}$$

which is known as ***Bayes's theorem***. The denominator $\Pr\{B\}$ is a normalizing constant that we can reexpress as follows. Since $B = (B \cap A) \cup (B \cap \overline{A})$ and $B \cap A$ and $B \cap \overline{A}$ are mutually exclusive events,

$$\begin{aligned}
\Pr\{B\} &= \Pr\{B \cap A\} + \Pr\{B \cap \overline{A}\} \\
&= \Pr\{A\}\Pr\{B \mid A\} + \Pr\{\overline{A}\}\Pr\{B \mid \overline{A}\}\,.
\end{aligned}$$

Substituting into equation (C.17), we obtain an equivalent form of Bayes's theorem:

$$\Pr\{A \mid B\} = \frac{\Pr\{A\}\Pr\{B \mid A\}}{\Pr\{A\}\Pr\{B \mid A\} + \Pr\{\overline{A}\}\Pr\{B \mid \overline{A}\}}\,.$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Bayes's theorem can simplify the computing of conditional probabilities. For example, suppose that we have a fair coin and a biased coin that always comes up heads. We run an experiment consisting of three independent events: one of the two coins is chosen at random, the coin is flipped once, and then it is flipped again. Suppose that the chosen coin comes up heads both times. What is the probability that it is biased?

We solve this problem using Bayes's theorem. Let A be the event that the biased coin is chosen, and let B be the event that the coin comes up heads both times. We wish to determine $\Pr\{A \mid B\}$. We have $\Pr\{A\} = 1/2$, $\Pr\{B \mid A\} = 1$, $\Pr\{\overline{A}\} = 1/2$, and $\Pr\{B \mid \overline{A}\} = 1/4$; hence,

$$\begin{aligned} \Pr\{A \mid B\} &= \frac{(1/2) \cdot 1}{(1/2) \cdot 1 + (1/2) \cdot (1/4)} \\ &= 4/5 \, . \end{aligned}$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

C.3 Discrete random variables

A ***(discrete) random variable*** X is a function from a finite or countably infinite sample space S to the real numbers. It associates a real number with each possible

outcome of an experiment, which allows us to work with the probability distribution induced on the resulting set of numbers. Random variables can also be defined for uncountably infinite sample spaces, but they raise technical issues that are unnecessary to address for our purposes. Henceforth, we shall assume that random variables are discrete.

For a random variable X and a real number x, we define the event $X = x$ to be $\{s \in S : X(s) = x\}$; thus,

$$\Pr\{X = x\} = \sum_{\{s \in S : X(s) = x\}} \Pr\{s\} \ .$$

The function

$$f(x) = \Pr\{X = x\}$$

is the ***probability density function*** of the random variable X. From the probability axioms, $\Pr\{X = x\} \geq 0$ and $\sum_x \Pr\{X = x\} = 1$.

As an example, consider the experiment of rolling a pair of ordinary, 6-sided dice. There are 36 possible elementary events in the sample space. We assume that the probability distribution is uniform, so that each elementary event $s \in S$ is equally likely: $\Pr\{s\} = 1/36$. Define the random variable X to be the *maximum* of the two values showing on the dice. We have $\Pr\{X = 3\} = 5/36$, since X assigns a value of 3 to 5 of the 36 possible elementary events, namely, (1, 3), (2, 3), (3, 3), (3, 2), and (3, 1).

It is common for several random variables to be defined on the same sample space. If X and Y are random variables, the function

$$f(x, y) = \Pr\{X = x \text{ and } Y = y\}$$

is the ***joint probability density function*** of X and Y. For a fixed value y,

$$\Pr\{Y = y\} = \sum_x \Pr\{X = x \text{ and } Y = y\} \ ,$$

and similarly, for a fixed value x,

$$\Pr\{X = x\} = \sum_y \Pr\{X = x \text{ and } Y = y\} \ .$$

Using the definition (C.14) of conditional probability, we have

$$\Pr\{X = x \mid Y = y\} = \frac{\Pr\{X = x \text{ and } Y = y\}}{\Pr\{Y = y\}} \ .$$

We define two random variables X and Y to be ***independent*** if for all x and y, the events $X = x$ and $Y = y$ are independent or, equivalently, if for all x and y, we have $\Pr\{X = x \text{ and } Y = y\} = \Pr\{X = x\} \Pr\{Y = y\}$.

Given a set of random variables defined over the same sample space, one can define new random variables as sums, products, or other functions of the original variables.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Expected value of a random variable

The simplest and most useful summary of the distribution of a random variable is the "average" of the values it takes on. The ***expected value*** (or, synonymously, ***expectation*** or ***mean***) of a discrete random variable X is

$$\mathrm{E}[X] = \sum_x x \Pr\{X = x\} , \tag{C.19}$$

which is well defined if the sum is finite or converges absolutely. Sometimes the expectation of X is denoted by μ_X or, when the random variable is apparent from context, simply by μ.

Consider a game in which you flip two fair coins. You earn \$3 for each head but lose \$2 for each tail. The expected value of the random variable X representing your earnings is

$$\begin{aligned}\mathrm{E}[X] &= 6 \cdot \Pr\{2 \text{ H's}\} + 1 \cdot \Pr\{1 \text{ H}, 1 \text{ T}\} - 4 \cdot \Pr\{2 \text{ T's}\} \\ &= 6(1/4) + 1(1/2) - 4(1/4) \\ &= 1 .\end{aligned}$$

The expectation of the sum of two random variables is the sum of their expectations, that is,

$$\mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y] , \tag{C.20}$$

whenever $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ are defined. We call this property ***linearity of expectation***, and it holds even if X and Y are not independent. It also extends to finite and absolutely convergent summations of expectations. Linearity of expectation is the key property that enables us to perform probabilistic analyses by using indicator random variables (see Section 5.2).

If X is any random variable, any function $g(x)$ defines a new random variable $g(X)$. If the expectation of $g(X)$ is defined, then

$$\mathrm{E}[g(X)] = \sum_x g(x) \Pr\{X = x\} .$$

Letting $g(x) = ax$, we have for any constant a,

$$\mathrm{E}[aX] = a\mathrm{E}[X] . \tag{C.21}$$

Consequently, expectations are linear: for any two random variables X and Y and any constant a,

$$\mathrm{E}[aX + Y] = a\mathrm{E}[X] + \mathrm{E}[Y] . \tag{C.22}$$

When two random variables X and Y are independent and each has a defined expectation,

$$\mathrm{E}[XY] = \sum_x \sum_y xy \Pr\{X = x \text{ and } Y = y\}$$

$$\begin{aligned} &= \sum_x \sum_y x y \Pr\{X = x\} \Pr\{Y = y\} \\ &= \left(\sum_x x \Pr\{X = x\}\right)\left(\sum_y y \Pr\{Y = y\}\right) \\ &= \mathrm{E}[X]\,\mathrm{E}[Y] \ . \end{aligned}$$

In general, when n random variables $X_1, X_2, \ldots, X_n$ are mutually independent,

$$\mathrm{E}[X_1 X_2 \cdots X_n] = \mathrm{E}[X_1]\,\mathrm{E}[X_2] \cdots \mathrm{E}[X_n] \ . \tag{C.23}$$

When a random variable X takes on values from the set of natural numbers $\mathbf{N} = \{0, 1, 2, \ldots\}$, there is a nice formula for its expectation:

$$\begin{aligned} \mathrm{E}[X] &= \sum_{i=0}^{\infty} i \Pr\{X = i\} \\ &= \sum_{i=0}^{\infty} i(\Pr\{X \geq i\} - \Pr\{X \geq i+1\}) \\ &= \sum_{i=1}^{\infty} \Pr\{X \geq i\} \ , \end{aligned} \tag{C.24}$$

since each term $\Pr\{X \geq i\}$ is added in i times and subtracted out $i-1$ times (except $\Pr\{X \geq 0\}$, which is added in 0 times and not subtracted out at all).

When we apply a convex function $f(x)$ to a random variable X, ***Jensen's inequality*** gives us

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X]) \ , \tag{C.25}$$

provided that the expectations exist and are finite. (A function $f(x)$ is ***convex*** if for all x and y and for all $0 \leq \lambda \leq 1$, we have $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$.)

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

C.4 The geometric and binomial distributions

A coin flip is an instance of a ***Bernoulli trial***, which is defined as an experiment with only two possible outcomes: ***success***, which occurs with probability p, and ***failure***, which occurs with probability $q = 1 - p$. When we speak of ***Bernoulli trials*** collectively, we mean that the trials are mutually independent and, unless we specifically say otherwise, that each has the same probability p for success. Two important distributions arise from Bernoulli trials: the geometric distribution and the binomial distribution.

The geometric distribution

Suppose we have a sequence of Bernoulli trials, each with a probability p of success and a probability $q = 1 - p$ of failure. How many trials occur before we obtain a success? Let the random variable X be the number of trials needed to obtain a success. Then X has values in the range $\{1, 2, \ldots\}$, and for $k \geq 1$,

$$\Pr\{X = k\} = q^{k-1}p \ , \tag{C.30}$$

since we have $k - 1$ failures before the one success. A probability distribution satisfying equation (C.30) is said to be a ***geometric distribution***. Figure C.1 illustrates such a distribution.

Assuming that $q < 1$, the expectation of a geometric distribution can be calculated using identity (A.8):

$$\begin{aligned} \mathrm{E}[X] &= \sum_{k=1}^{\infty} kq^{k-1}p \\ &= \frac{p}{q}\sum_{k=0}^{\infty} kq^{k} \\ &= \frac{p}{q}\cdot\frac{q}{(1-q)^2} \\ &= 1/p \ . \end{aligned} \tag{C.31}$$

Thus, on average, it takes $1/p$ trials before we obtain a success, an intuitive result.

As an example, suppose we repeatedly roll two dice until we obtain either a seven or an eleven. Of the 36 possible outcomes, 6 yield a seven and 2 yield an eleven. Thus, the probability of success is $p = 8/36 = 2/9$, and we must roll $1/p = 9/2 = 4.5$ times on average to obtain a seven or eleven.

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

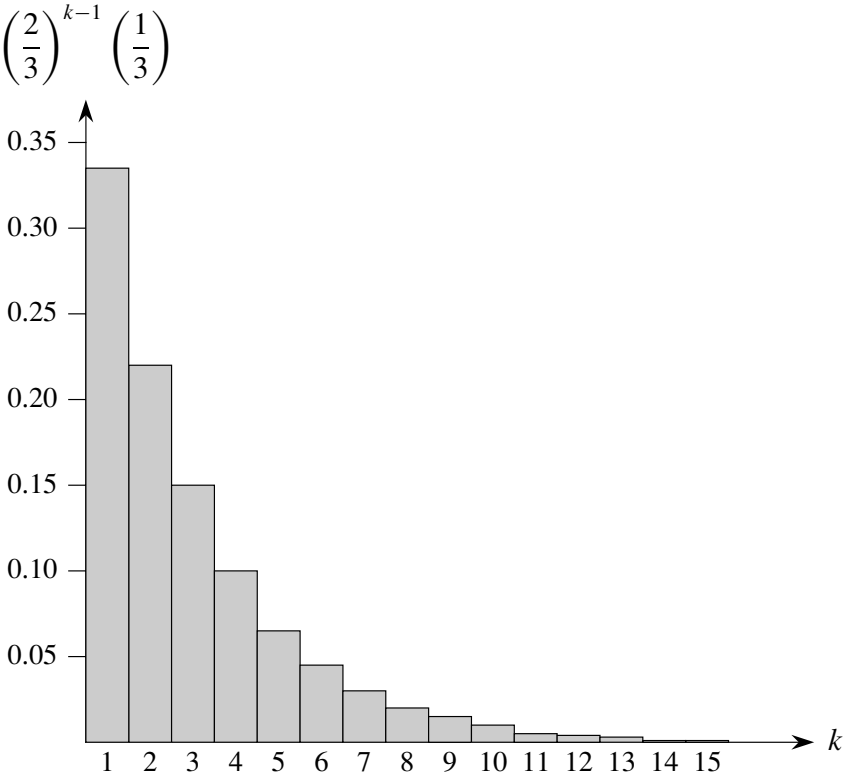

Figure C.1 A geometric distribution with probability $p = 1/3$ of success and a probability $q = 1 - p$ of failure. The expectation of the distribution is $1/p = 3$.

The binomial distribution

How many successes occur during n Bernoulli trials, where a success occurs with probability p and a failure with probability $q = 1 - p$? Define the random variable X to be the number of successes in n trials. Then X has values in the range $\{0, 1, \ldots, n\}$, and for $k = 0, \ldots, n$,

$$\Pr\{X = k\} = \binom{n}{k} p^k q^{n-k} , \tag{C.33}$$

since there are $\binom{n}{k}$ ways to pick which k of the n trials are successes, and the probability that each occurs is $p^k q^{n-k}$. A probability distribution satisfying equation (C.33) is said to be a ***binomial distribution***. For convenience, we define the family of binomial distributions using the notation

$$b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} . \tag{C.34}$$

Figure C.2 illustrates a binomial distribution. The name "binomial" comes from the fact that (C.33) is the kth term of the expansion of $(p + q)^n$. Consequently, since $p + q = 1$,

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

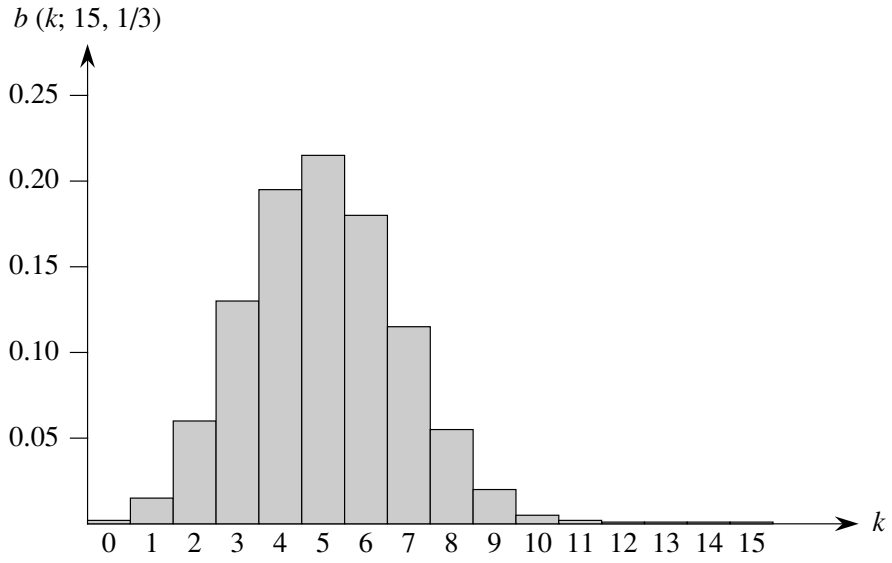

Figure C.2 The binomial distribution $b(k; 15, 1/3)$ resulting from $n = 15$ Bernoulli trials, each with probability $p = 1/3$ of success. The expectation of the distribution is $np = 5$.

$$\sum_{k=0}^{n} b(k; n, p) = 1 , \tag{C.35}$$

as is required by axiom 2 of the probability axioms.

We can compute the expectation of a random variable having a binomial distribution from equations (C.8) and (C.35). Let X be a random variable that follows the binomial distribution $b(k; n, p)$, and let $q = 1 - p$. By the definition of expectation, we have

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{n} k \Pr\{X = k\} \\
&= \sum_{k=0}^{n} k\, b(k; n, p) \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{(n-1)-k}
\end{aligned}$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

$$
\begin{aligned}
&= np \sum_{k=0}^{n-1} b(k; n-1, p) \\
&= np \, . \qquad \text{(C.36)}
\end{aligned}
$$

By using the linearity of expectation, we can obtain the same result with substantially less algebra. Let X_i be the random variable describing the number of successes in the ith trial. Then $\mathrm{E}[X_i] = p \cdot 1 + q \cdot 0 = p$, and by linearity of expectation (equation (C.20)), the expected number of successes for n trials is

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \sum_{i=1}^{n} p \\
&= np \, . \qquad \text{(C.37)}
\end{aligned}
$$

Copyright © 2001. MIT Press.
All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

Introduction to Python 3

We conclude with an introduction to Python 3.

- Look over and try some examples from `https://docs.python.org/3/tutorial/`
- In particular, it is useful to peruse and try out, from `https://docs.python.org/3/library/index.html#library-index`, the following:
  - Built-in Functions
  - Build-in Constants
  - Built-in Types
    - The principal built-in types are: numerics, sequences, mappings, classes, instances and exceptions. Of these, you will most likely not have to worry about classes, instances and exceptions. The code you will need for this course will be quite straightforward.
- You can of course install Python 3 on your personal device. Also, Python 3 is installed on eceUbuntu. As a student in ECE, you should already have an account on eceUbuntu. If you are on campus, you should be able to simply ssh eceUbuntu. If you are off campus, you can either install the campus VPN:

  `https://uwaterloo.ca/information-systems-technology/services/virtual-private-network-vpn`

  Or first ssh ecelinux4.uwaterloo.ca and then immediately, ssh -X eceUbuntu as the instructions say.
- It is useful to do the following. Better yet, add it to your `.bashrc` file.

  ```
  [alice@ecetesla2 ~]$ alias python='/usr/bin/python3'
  ```
- Example python code on Learn:
  - ask.py, and,
  - romandecimal.py, romandecimalsolution.py, tester-rnstringtodec.py