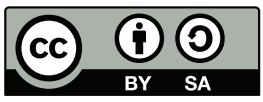
Discrete Structures for Computer Science: Counting, Recursion, and Probability

Michiel Smid

School of Computer Science
Carleton University
Ottawa
Canada
michiel@scs.carleton.ca

December 12, 2014



Contents

Preface **vi**

1 Introduction **1**
1.1 Ramsey Theory 1
1.2 Sperner's Theorem 4
1.3 The Quick-Sort Algorithm 5

2 Mathematical Preliminaries **9**
2.1 Basic Concepts 9
2.2 Proof Techniques 11
2.2.1 Direct proofs 13
2.2.2 Constructive proofs 14
2.2.3 Nonconstructive proofs 14
2.2.4 Proofs by contradiction 15
2.2.5 Proofs by induction 16
2.2.6 More examples of proofs 18
2.3 Asymptotic Notation 20
2.4 Logarithms 21
2.5 Exercises 23

3 Counting **25**
3.1 The Product Rule 25
3.1.1 Counting Bitstrings of Length n 26
3.1.2 Counting Functions 26
3.1.3 Placing Books on Shelves 29
3.2 The Bijection Rule 30
3.3 The Complement Rule 32
3.4 The Sum Rule 33

3.5 The Principle of Inclusion and Exclusion 34
3.6 Permutations and Binomial Coefficients 36
3.6.1 Some Examples 38
3.6.2 Newton's Binomial Theorem 39
3.7 Combinatorial Proofs 41
3.8 Pascal's Triangle 45
3.9 More Counting Problems 48
3.9.1 Reordering the Letters of a Word 48
3.9.2 Counting Solutions of Linear Equations 49
3.10 The Pigeonhole Principle 53
3.10.1 India Pale Ale 53
3.10.2 Sequences Containing Divisible Numbers 54
3.10.3 Long Sequences Contain Long Monotone Subsequences 55
3.10.4 There are Infinitely Many Primes 56
3.11 Exercises 57

4 Recursion 65
4.1 Recursive Functions 65
4.2 Fibonacci Numbers 67
4.2.1 Counting 00-Free Bitstrings 69
4.3 A Recursively Defined Set 70
4.4 A Gossip Problem 72
4.5 The Merge-Sort Algorithm 76
4.5.1 Correctness of Algorithm MergeSort 77
4.5.2 Running Time of Algorithm MergeSort 78
4.6 Counting Regions when Cutting a Circle 81
4.6.1 A Polynomial Upper Bound on R_n 82
4.6.2 A Recurrence Relation for R_n 85
4.6.3 Simplifying the Recurrence Relation 90
4.6.4 Solving the Recurrence Relation 91
4.7 Exercises 92

5 Discrete Probability 105
5.1 Anonymous Broadcasting 105
5.2 Probability Spaces 110
5.2.1 Examples 111
5.3 Basic Rules of Probability 114
5.4 Uniform Probability Spaces 119

5.4.1 The Probability of Getting a Full House 120
5.5 The Birthday Paradox 121
5.5.1 Throwing Balls into Boxes 123
5.6 The Big Box Problem 125
5.6.1 The Probability of Finding the Big Box 126
5.7 The Monty Hall Problem 128
5.8 Conditional Probability 129
5.8.1 Anil's Children 131
5.8.2 Rolling a Die 131
5.9 The Law of Total Probability 133
5.9.1 Flipping a Coin and Rolling Dice 135
5.10 Independent Events 137
5.10.1 Rolling Two Dice 137
5.10.2 A Basic Property of Independent Events 138
5.10.3 Pairwise and Mutually Independent Events 139
5.11 Describing Events by Logical Propositions 141
5.11.1 Flipping a Coin and Rolling a Die 141
5.11.2 Flipping Coins 142
5.11.3 The Probability of a Circuit Failing 143
5.12 Choosing a Random Element in a Linked List 144
5.13 Long Runs in Random Bitstrings 146
5.14 Infinite Probability Spaces 151
5.14.1 Infinite Series 152
5.14.2 Who Flips the First Heads 154
5.14.3 Who Flips the Second Heads 156
5.15 Exercises 158

6 Random Variables and Expectation 171
6.1 Random Variables 171
6.1.1 Flipping Three Coins 171
6.1.2 Random Variables and Events 173
6.2 Independent Random Variables 175
6.3 Distribution Functions 177
6.4 Expected Values 179
6.4.1 Some Examples 179
6.4.2 An Alternative Formula for the Expected Value 181
6.5 Linearity of Expectation 184
6.6 The Geometric Distribution 187

6.6.1 Determining the Expected Value . . . 187
6.7 The Binomial Distribution . . . 189
6.7.1 Determining the Expected Value . . . 190
6.7.2 Using the Linearity of Expectation . . . 192
6.8 Indicator Random Variables . . . 193
6.8.1 Runs in Random Bitstrings . . . 194
6.8.2 Largest Elements in Prefixes of Random Permutations 196
6.8.3 Estimating the Harmonic Number . . . 198
6.9 The Insertion-Sort Algorithm . . . 201
6.10 The Quick-Sort Algorithm . . . 203
6.11 Skip Lists . . . 206
6.11.1 Algorithm SEARCH . . . 208
6.11.2 Algorithms INSERT and DELETE . . . 209
6.11.3 Analysis of Skip Lists . . . 211
6.12 Exercises . . . 218

7 The Probabilistic Method 231
7.1 Large Bipartite Subgraphs . . . 231
7.2 Ramsey Theory . . . 233
7.3 Sperner's Theorem . . . 236
7.4 Planar Graphs and the Crossing Lemma . . . 239
7.4.1 Planar Graphs . . . 239
7.4.2 The Crossing Number of a Graph . . . 243
7.5 Exercises . . . 248

Preface

This is a free textbook for an undergraduate course on Discrete Structures for Computer Science students, which I have been teaching at Carleton University since the fall term of 2013. The material is offered as the second-year course COMP 2804 (Discrete Structures II). Students are assumed to have taken COMP 1805 (Discrete Structures I), which covers mathematical reasoning, basic proof techniques, sets, functions, relations, basic graph theory, asymptotic notation, and countability.

During a 12-week term with three hours of classes per week, I cover most of the material in this book, except for Chapter 2, which has been included so that students can review the material from COMP 1805.

Chapter 2 is largely taken from the free textbook *Introduction to Theory of Computation* by Anil Maheshwari and Michiel Smid, which is available at `http://cg.scs.carleton.ca/~michiel/TheoryOfComputation/`

Please let me know if you find errors, typos, simpler proofs, comments, omissions, or if you think that some parts of the book "need improvement".

Chapter 1

Introduction

In this chapter, we introduce some problems that will be solved later in this book. Along the way, we recall some notions from discrete mathematics that you are assumed to be familiar with. These notions are reviewed in more detail in Chapter 2.

1.1 Ramsey Theory

Ramsey Theory studies problems of the following form: How many elements of a given type must there be so that we can guarantee that some property holds? In this section, we consider the case when the elements are people and the property is "there is a large group of friends or there is a large group of strangers".

Theorem 1.1.1 *In any group of six people, there are*

- *three friends, i.e., three people who know each other,*
- *or three strangers, i.e., three people, none of which knows the other two.*

In order to prove this theorem, we denote the six people by $P_1, P_2, \ldots, P_6$. Any two people P_i and P_j are either *friends* or *strangers*. We define the complete graph $G = (V, E)$ with vertex set

$$V = \{P_i : 1 \leq i \leq 6\}$$

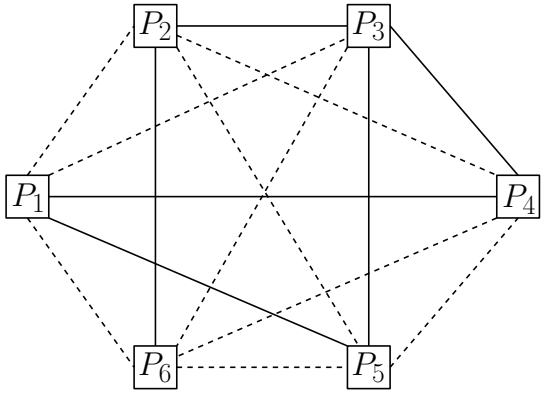
and edge set

$$E = \{P_iP_j : 1 \leq i < j \leq 6\}.$$

Observe that $|V| = 6$ and $|E| = 15$. We draw each edge P_iP_j as a straight-line segment according to the following rule:

- If P_i and P_j are friends, then the edge P_iP_j is *solid*.
- If P_i and P_j are strangers, then the edge P_iP_j is *dashed*.

In the example below, P_3 and P_5 are friends, whereas P_1 and P_3 are strangers.



Observe that a group of three friends corresponds to a solid triangle in the graph G, whereas a group of three strangers corresponds to a dashed triangle. In the example above, there is no solid triangle and, thus, there is no group of three friends. Since the triangle $P_2P_4P_5$ is dashed, there is a group of three strangers.

Using this terminology, Theorem 1.1.1 is equivalent to the following:

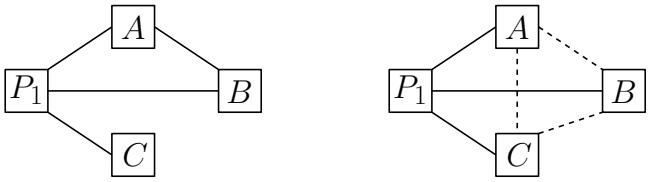
Theorem 1.1.2 *Consider a complete graph on six vertices, in which each edge is either solid or dashed. Then there is a solid triangle or a dashed triangle.*

Proof. As before, we denote the vertices by $P_1, \ldots, P_6$. Consider the five edges that are incident on P_1. Using a proof by contradiction, it can easily be shown that one of the following two claims must hold:

- At least three of these five edges are solid.
- At least three of these five edges are dashed.

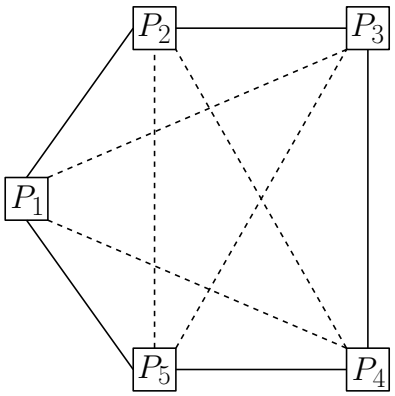
We may assume without loss of generality that the first claim holds. Consider three edges incident on P_1 that are solid and denote them by P_1A, P_1B, and P_1C.

If at least one of the edges AB, AC, and BC is solid, then there is a solid triangle. In the left figure below, AB is solid and we obtain the solid triangle P_1AB.



Otherwise, all edges AB, AC, and BC are dashed, in which case we obtain the dashed triangle ABC; see the right figure above. ∎

The example below shows an example of a complete graph with five vertices without any solid triangle and without any dashed triangle. Thus, Theorem 1.1.2 does not hold for complete graphs with five vertices. Equivalently, Theorem 1.1.1 does not hold for groups of five people.



What about larger groups of friends/strangers? Let $k \geq 3$ be an integer. The following theorem states that even if we take $\lfloor 2^{k/2} \rfloor$ people, we are not guaranteed that there is a group of k friends or a group of k strangers.

A *k-clique* in a graph is a set of k vertices, any two of which are connected by an edge. For example, a 3-clique is a triangle.

Theorem 1.1.3 *Let $k \geq 3$ and $n \leq \lfloor 2^{k/2} \rfloor$ be integers. There exists a complete graph with n vertices, in which each edge is either solid or dashed, such that this graph does not contain a solid k-clique and does not contain a dashed k-clique.*

We will prove this theorem in Section 7.2 using elementary counting techniques and probability theory. This probably sounds surprising to you, because Theorem 1.1.3 does not have anything to do with probability. In fact, in Section 7.2, we will prove the following claim: Take $k = 20$ and $n = 1024$. If you go to the ByWard Market in downtown Ottawa and take a random group of n people, then with very high probability, this group does not contain a subgroup of k friends *and* does not contain a subgroup of k strangers. In other words, with very high probability, *every* subgroup of k people contains two friends *and* two strangers.

1.2 Sperner's Theorem

Consider a set S with five elements, say, $S = \{1, 2, 3, 4, 5\}$. Let $S_1, S_2, \ldots, S_m$ be a sequence of m subsets of S, such that for all i and j with $i \neq j$,

$$S_i \not\subseteq S_j \text{ and } S_j \not\subseteq S_i,$$

i.e., S_i is not a subset of S_j and S_j is not a subset of S_i. How large can m be? The following example shows that m can be as large as 10:

$$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}, \{4,5\}$$

(Note that these are all 10 subsets of S having size two.) Can there be such a sequence of more than 10 subsets? The following theorem states that the answer is "no".

Theorem 1.2.1 (Sperner) *Let $n \geq 1$ be an integer and let S be a set with n elements. Let $S_1, S_2, \ldots, S_m$ be a sequence of m subsets of S, such that for all i and j with $i \neq j$,*

$$S_i \not\subseteq S_j \text{ and } S_j \not\subseteq S_i.$$

Then

$$m \leq \binom{n}{\lfloor n/2 \rfloor}.$$

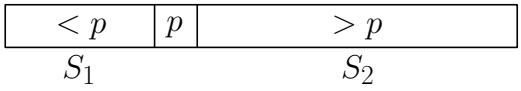
The right-hand side of the last line is a binomial coefficient, which we will define in Section 3.6. Its value is equal to the number of subsets of S having size $\lfloor n/2 \rfloor$. Observe that these subsets satisfy the property in Theorem 1.2.1.

We will prove Theorem 1.2.1 in Section 7.3 using elementary counting techniques and probability theory. Again, this probably sounds surprising to you, because Theorem 1.2.1 does not have anything to do with probability.

1.3 The Quick-Sort Algorithm

You are probably familiar with the QuickSort algorithm. This algorithm sorts any sequence S of $n \geq 0$ pairwise distinct numbers in the following way:

- If $n = 0$ of $n = 1$, then there is nothing to do.
- If $n \geq 2$, then the algorithm picks one of the numbers in S, let us call it p (which stands for *pivot*), scans the sequence S, and splits it into three subsequences: one subsequence S_1 contains all elements in S that are less than p, one subsequence only consists of the element p, and the third subsequence S_2 contains all elements in S that are larger than p; see the figure below.



  The algorithm *recursively* runs QuickSort on the subsequence S_1, after which it, again recursively, runs QuickSort on the subsequence S_2.

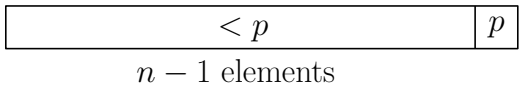
Running QuickSort recursively on the subsequence S_1 means that we first check if S_1 has size at most one; if this is the case, nothing needs to be done, because S_1 is sorted already. If S_1 has size at least two, then we choose a pivot p_1 in S_1, use p_1 to split S_1 into three subsequences, and recursively run QuickSort on the subsequence of S_1 consisting of all elements that are less than p_1, and finally run QuickSort on the subsequence of S_1 consisting of all elements that are larger than p_1. (We will see recursive algorithms in more detail in Chapter 4.)

Algorithm QuickSort correctly sorts any sequence of numbers, no matter how we choose the pivot element. It turns out, however, that the running

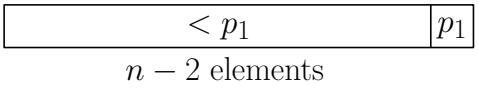
time of the algorithm heavily depends on the pivots that are chosen in the recursive calls.

For example, assume that in each (recursive) call to the algorithm, the pivot happens to be the largest element in the sequence. Then, in each call, the subsequence of elements that are larger than the pivot is empty. Let us see what happens in this case:

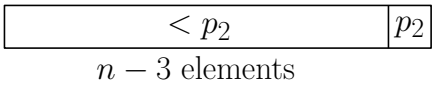
- We start with a sequence S of size n. The first pivot p is the largest element in S. Thus, using the notation given above, the subsequence S_1 contains $n-1$ elements (all elements of S except for p), whereas the subsequence S_2 is empty. Computing these subsequences can be done in n "steps", after which we are in the following situation:



- We now run QuickSort on a sequence of $n-1$ elements. Again, the pivot p_1 is the largest element. In $n-1$ steps, we obtain a subsequence of $n-2$ elements that are less than p_1, and an empty subsequence of elements that are larger than p_1; see the figure below.



- Next we run QuickSort on a sequence of $n-2$ elements. As before, the pivot p_2 is the largest element. In $n-2$ steps, we obtain a subsequence of $n-3$ elements that are less than p_2, and an empty subsequence of elements that are larger than p_2; see the figure below.



You probably see the pattern. The total running time of the algorithm is proportional to

$$n + (n-1) + (n-2) + (n-3) + \cdots + 3 + 2 + 1,$$

which, by Theorem 2.2.10, is equal to

$$\frac{1}{2}\,n(n+1) = \frac{1}{2}\,n^2 + \frac{1}{2}\,n,$$

which, using the Big-Theta notation (see Section 2.3) is $\Theta(n^2)$, i.e., quadratic in n. It can be shown that this is, in fact, the worst-case running time of QuickSort.

What would be a good choice for the pivot elements? Intuitively, a pivot is good if the sequences S_1 and S_2 have (roughly) the same size. Thus, after the first call, we are in the following situation:

$< p$	p	$> p$
$(n-1)/2$		$(n-1)/2$

In Section 4.5, we will prove that, if this happens in each recursive call, the running time of the algorithm is only $O(n \log n)$. Obviously, it is not clear at all how we can guarantee that we always choose a good pivot. It turns out that there is an easy strategy: In each call, choose the pivot *randomly*! That is, among all elements involved in the recursive call, pick one uniformly at random so that each element has the same probability of being chosen. In Section 6.10, we will prove that this leads to an *expected* running time of $O(n \log n)$.

Chapter 2

Mathematical Preliminaries

2.1 Basic Concepts

Throughout this book, we will assume that you know the following mathematical concepts:

1. A *set* is a collection of well-defined objects. Examples are (i) the set of all Dutch Olympic Gold Medallists, (ii) the set of all pubs in Ottawa, and (iii) the set of all even natural numbers.

2. The set of *natural numbers* is $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$.

3. The set of *integers* is $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

4. The set of *rational numbers* is $\mathbb{Q} = \{m/n : m \in \mathbb{Z}, n \in \mathbb{Z}, n \neq 0\}$.

5. The set of *real numbers* is denoted by $\mathbb{R}$.

6. The *empty set* is the set that does not contain any element. This set is denoted by $\emptyset$.

7. If A and B are sets, then A is a *subset* of B, written as $A \subseteq B$, if every element of A is also an element of B. For example, the set of even natural numbers is a subset of the set of all natural numbers. Every set A is a subset of itself, i.e., $A \subseteq A$. The empty set is a subset of every set A, i.e., $\emptyset \subseteq A$. We say that A is a *proper subset* of B, written as $A \subset B$, if $A \subseteq B$ and $A \neq B$.

8. If B is a set, then the *power set* $\mathcal{P}(B)$ of B is defined to be the set of all subsets of B:
$$\mathcal{P}(B) = \{A : A \subseteq B\}.$$
Observe that $\emptyset \in \mathcal{P}(B)$ and $B \in \mathcal{P}(B)$.

9. If A and B are two sets, then

   (a) their *union* is defined as
   $$A \cup B = \{x : x \in A \text{ or } x \in B\},$$

   (b) their *intersection* is defined as
   $$A \cap B = \{x : x \in A \text{ and } x \in B\},$$

   (c) their *difference* is defined as
   $$A \setminus B = \{x : x \in A \text{ and } x \notin B\},$$

   (d) the *Cartesian product* of A and B is defined as
   $$A \times B = \{(x, y) : x \in A \text{ and } y \in B\},$$

   (e) the *complement* of A is defined as
   $$\overline{A} = \{x : x \notin A\}.$$

10. A *binary relation* on two sets A and B is a subset of $A \times B$.

11. A *function* f from A to B, denoted by $f : A \to B$, is a binary relation R, having the property that for each element $a \in A$, there is exactly one ordered pair in R, whose first component is a. We will also say that $f(a) = b$, or f maps a to b, or the image of a under f is b. The set A is called the *domain* of f, and the set
$$\{b \in B : \text{ there is an } a \in A \text{ with } f(a) = b\}$$
is called the *range* of f.

12. A function $f : A \to B$ is *one-to-one* (or *injective*), if for any two distinct elements a and a' in A, we have $f(a) \neq f(a')$. The function f is *onto* (or *surjective*), if for each element $b \in B$, there exists an element $a \in A$, such that $f(a) = b$; in other words, the range of f is equal to the set B. A function f is a *bijection*, if f is both injective and surjective.

13. A set A is *countable*, if A is finite or there is a bijection $f : \mathbb{N} \to A$. The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countable, whereas $\mathbb{R}$ is not.

14. A *graph* $G = (V, E)$ is a pair consisting of a set V, whose elements are called *vertices*, and a set E, where each element of E is a pair of distinct vertices. The elements of E are called *edges*.

15. The *Boolean values* are 1 and 0, that represent *true* and *false*, respectively. The basic Boolean operations include

    (a) negation (or NOT), represented by $\neg$,

    (b) conjunction (or AND), represented by $\wedge$,

    (c) disjunction (or OR), represented by $\vee$,

    (d) exclusive-or (or XOR), represented by $\oplus$,

    (e) equivalence, represented by $\leftrightarrow$ or $\Leftrightarrow$,

    (f) implication, represented by $\rightarrow$ or $\Rightarrow$.

    The following table explains the meanings of these operations.

NOT	AND	OR	XOR	equivalence	implication
$\neg 0 = 1$	$0 \wedge 0 = 0$	$0 \vee 0 = 0$	$0 \oplus 0 = 0$	$0 \leftrightarrow 0 = 1$	$0 \rightarrow 0 = 1$
$\neg 1 = 0$	$0 \wedge 1 = 0$	$0 \vee 1 = 1$	$0 \oplus 1 = 1$	$0 \leftrightarrow 1 = 0$	$0 \rightarrow 1 = 1$
	$1 \wedge 0 = 0$	$1 \vee 0 = 1$	$1 \oplus 0 = 1$	$1 \leftrightarrow 0 = 0$	$1 \rightarrow 0 = 0$
	$1 \wedge 1 = 1$	$1 \vee 1 = 1$	$1 \oplus 1 = 0$	$1 \leftrightarrow 1 = 1$	$1 \rightarrow 1 = 1$

2.2 Proof Techniques

> A proof is a proof. What kind of a proof? It's a proof. A proof is a proof. And when you have a good proof, it's because it's proven.
>
> — Jean Chrétien, Prime Minister of Canada (1993–2003)

In mathematics, a theorem is a statement that is true. A proof is a sequence of mathematical statements that form an argument to show that a theorem is true. The statements in the proof of a theorem include axioms (assumptions about the underlying mathematical structures), hypotheses of the theorem to be proved, and previously proved theorems. The main question is "How do we go about proving theorems?" This question is similar to the question of how to solve a given problem. Of course, the answer is that finding proofs, or solving problems, is not easy; otherwise life would be dull! There is no specified way of coming up with a proof, but there are some generic strategies that could be of help. In this section, we review some of these strategies, that will be sufficient for this course. The best way to get a feeling of how to come up with a proof is by solving a large number of problems. Here are some useful tips. (You may take a look at the book *How to Solve It*, by George Pólya).

1. Read and completely understand the statement of the theorem to be proved. Most often this is the hardest part.

2. Sometimes, theorems contain theorems inside them. For example, "Property A if and only if property B", requires showing two statements:

   (a) If property A is true, then property B is true ($A \Rightarrow B$).

   (b) If property B is true, then property A is true ($B \Rightarrow A$).

   Another example is the theorem "Set A equals set B." To prove this, we need to prove that $A \subseteq B$ and $B \subseteq A$. That is, we need to show that each element of set A is in set B, and each element of set B is in set A.

3. Try to work out a few simple cases of the theorem just to get a grip on it (i.e., crack a few simple cases first).

4. Try to write down the proof once you think you have it. This is to ensure the correctness of your proof. Often, mistakes are found at the time of writing.

5. Finding proofs takes time, we do not come prewired to produce proofs. Be patient, think, express and write clearly, and try to be precise as much as possible.

In the next sections, we will go through some of the proof strategies.

2.2.1 Direct proofs

As the name suggests, in a direct proof of a theorem, we just approach the theorem directly.

Theorem 2.2.1 *If n is an odd positive integer, then n^2 is odd as well.*

Proof. An odd positive integer n can be written as $n = 2k + 1$, for some integer $k \geq 0$. Then

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1.$$

Since $2(2k^2 + 2k)$ is even, and "even plus one is odd", we can conclude that n^2 is odd. ∎

For a graph $G = (V, E)$, the *degree* of a vertex v, denoted by $deg(v)$, is defined to be the number of edges that are incident on v.

Theorem 2.2.2 *Let $G = (V, E)$ be a graph. Then the sum of the degrees of all vertices is an even integer, i.e.,*

$$\sum_{v \in V} deg(v)$$

is even.

Proof. If you do not see the meaning of this statement, then first try it out for a few graphs. The reason why the statement holds is very simple: Each edge contributes 2 to the summation (because an edge is incident on exactly two distinct vertices). ∎

Actually, the proof above proves the following theorem.

Theorem 2.2.3 *Let $G = (V, E)$ be a graph. Then the sum of the degrees of all vertices is equal to twice the number of edges, i.e.,*

$$\sum_{v \in V} deg(v) = 2|E|.$$

2.2.2 Constructive proofs

This technique not only shows the existence of a certain object, it actually gives a method of creating it:

Theorem 2.2.4 *There exists an object with property $\mathcal{P}$.*

Proof. Here is the object: [...]
And here is the proof that the object satisfies property $\mathcal{P}$: [...] ∎

A graph is called 3-*regular*, if each vertex has degree three. We prove the following theorem using a constructive proof.

Theorem 2.2.5 *For every even integer $n \geq 4$, there exists a 3-regular graph with n vertices.*

Proof. Define

$$V = \{0, 1, 2, \ldots, n-1\},$$

and

$$E = \{\{i, i+1\} : 0 \leq i \leq n-2\} \cup \{\{n-1, 0\}\} \cup \{\{i, i+n/2\} : 0 \leq i \leq n/2-1\}.$$

Then the graph $G = (V, E)$ is 3-regular.

Convince yourself that this graph is indeed 3-regular. It may help to draw the graph for, say, $n = 8$. ∎

2.2.3 Nonconstructive proofs

In a nonconstructive proof, we show that a certain object exists, without actually creating it. Here is an example of such a proof:

Theorem 2.2.6 *There exist irrational numbers x and y such that x^y is rational.*

Proof. There are two possible cases.

Case 1: $\sqrt{2}^{\sqrt{2}} \in \mathbb{Q}$.

In this case, we take $x = y = \sqrt{2}$. In Theorem 2.2.9 below, we will prove that $\sqrt{2}$ is irrational.

Case 2: $\sqrt{2}^{\sqrt{2}} \notin \mathbb{Q}$.

In this case, we take $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$. Since

$$x^y = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \sqrt{2}^2 = 2,$$

the claim in the theorem follows. ∎

Observe that this proof indeed proves the theorem, but it does not give an example of a pair of irrational numbers x and y such that x^y is rational.

2.2.4 Proofs by contradiction

This is how a proof by contradiction looks like:

Theorem 2.2.7 *Statement $\mathcal{S}$ is true.*

Proof. Assume that statement $\mathcal{S}$ is false. Then, derive a contradiction (such as $1 + 1 = 3$).

In other words, show that the statement "$\neg\mathcal{S} \Rightarrow$ *false*" is true. This is sufficient, because the contrapositive of the statement "$\neg\mathcal{S} \Rightarrow$ *false*" is the statement "*true* $\Rightarrow \mathcal{S}$". The latter logical formula is equivalent to $\mathcal{S}$, and that is what we wanted to show. ∎

Below, we give two examples of proofs by contradiction.

Theorem 2.2.8 *Let n be a positive integer. If n^2 is even, then n is even.*

Proof. We will prove the theorem by contradiction. So we assume that n^2 is even, but n is odd. Since n is odd, we know from Theorem 2.2.1 that n^2 is odd. This is a contradiction, because we assumed that n^2 is even. ∎

Theorem 2.2.9 *$\sqrt{2}$ is irrational, i.e., $\sqrt{2}$ cannot be written as a fraction of two integers.*

Proof. We will prove the theorem by contradiction. So we assume that $\sqrt{2}$ is rational. Then $\sqrt{2}$ can be written as a fraction of two integers $m \geq 1$ and $n \geq 1$, i.e., $\sqrt{2} = m/n$. We may assume that m and n are not both

even, because otherwise, we can get rid of the common factors. By squaring $\sqrt{2} = m/n$, we get $2n^2 = m^2$. This implies that m^2 is even. Then, by Theorem 2.2.8, m is even, which means that we can write m as $m = 2k$, for some positive integer k. It follows that $2n^2 = m^2 = 4k^2$, which implies that $n^2 = 2k^2$. Hence, n^2 is even. Again by Theorem 2.2.8, it follows that n is even.

We have shown that m and n are both even. But we know that m and n are *not* both even. Hence, we have a contradiction. Our assumption that $\sqrt{2}$ is rational is wrong. Thus, we can conclude that $\sqrt{2}$ is irrational. ∎

There is a nice discussion of this proof in the book *My Brain is Open: The Mathematical Journeys of Paul Erdős* by Bruce Schechter.

2.2.5 Proofs by induction

This is a very powerful and important technique for proving theorems.

For each positive integer n, let $P(n)$ be a mathematical statement that depends on n. Assume we wish to prove that $P(n)$ is true for all positive integers n. A proof by induction of such a statement is carried out as follows:

Basis: Prove that $P(1)$ is true.

Induction step: Prove that for all $n \geq 1$, the following holds: If $P(n)$ is true, then $P(n+1)$ is also true.

In the induction step, we choose an arbitrary integer $n \geq 1$ and assume that $P(n)$ is true; this is called the *induction hypothesis*. Then we prove that $P(n+1)$ is also true.

Theorem 2.2.10 *For all positive integers n, we have*

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. We start with the basis of the induction. If $n = 1$, then the left-hand side is equal to 1, and so is the right-hand side. So the theorem is true for $n = 1$.

For the induction step, let $n \geq 1$ and assume that the theorem is true for n, i.e., assume that

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

We have to prove that the theorem is true for $n+1$, i.e., we have to prove that

$$1+2+3+\cdots+(n+1)=\frac{(n+1)(n+2)}{2}.$$

Here is the proof:

$$\begin{aligned}1+2+3+\cdots+(n+1) &= \underbrace{1+2+3+\cdots+n}_{=\frac{n(n+1)}{2}}+(n+1)\\ &= \frac{n(n+1)}{2}+(n+1)\\ &= \frac{(n+1)(n+2)}{2}.\end{aligned}$$

∎

By the way, here is an alternative proof of the theorem above: Let $S = 1+2+3+\cdots+n$. Then,

S	$=$	1	$+$	2	$+$	3	$+$	$\cdots$	$+$	$(n-2)$	$+$	$(n-1)$	$+$	n
S	$=$	n	$+$	$(n-1)$	$+$	$(n-2)$	$+$	$\cdots$	$+$	3	$+$	2	$+$	1
$2S$	$=$	$(n+1)$	$+$	$(n+1)$	$+$	$(n+1)$	$+$	$\cdots$	$+$	$(n+1)$	$+$	$(n+1)$	$+$	$(n+1)$

Since there are n terms on the right-hand side, we have $2S = n(n+1)$. This implies that $S = n(n+1)/2$.

Theorem 2.2.11 *For every positive integer n, $a-b$ is a factor of a^n-b^n.*

Proof. A direct proof can be given by providing a factorization of a^n-b^n:

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

We now prove the theorem by induction. For the basis, let $n=1$. The claim in the theorem is "$a-b$ is a factor of $a-b$", which is obviously true.

Let $n \geq 1$ and assume that $a-b$ is a factor of a^n-b^n. We have to prove that $a-b$ is a factor of $a^{n+1}-b^{n+1}$. We have

$$a^{n+1}-b^{n+1}=a^{n+1}-a^nb+a^nb-b^{n+1}=a^n(a-b)+(a^n-b^n)b.$$

The first term on the right-hand side is divisible by $a-b$. By the induction hypothesis, the second term on the right-hand side is divisible by $a-b$ as well. Therefore, the entire right-hand side is divisible by $a-b$. Since the right-hand side is equal to $a^{n+1}-b^{n+1}$, it follows that $a-b$ is a factor of $a^{n+1}-b^{n+1}$. ∎

We now give an alternative proof of Theorem 2.2.3:

Theorem 2.2.12 *Let $G = (V, E)$ be a graph with m edges. Then the sum of the degrees of all vertices is equal to twice the number of edges, i.e.,*

$$\sum_{v \in V} deg(v) = 2m.$$

Proof. The proof is by induction on the number m of edges. For the basis of the induction, assume that $m = 0$. Then the graph G does not contain any edges and, therefore, $\sum_{v \in V} deg(v) = 0$. Thus, the theorem is true if $m = 0$.

Let $m \geq 0$ and assume that the theorem is true for every graph with m edges. Let G be an arbitrary graph with $m+1$ edges. We have to prove that $\sum_{v \in V} deg(v) = 2(m + 1)$.

Let $\{a, b\}$ be an arbitrary edge in G, and let G' be the graph obtained from G by removing the edge $\{a, b\}$. Since G' has m edges, we know from the induction hypothesis that the sum of the degrees of all vertices in G' is equal to $2m$. Using this, we obtain

$$\sum_{v \in G} deg(v) = \sum_{v \in G'} deg(v) + 2 = 2m + 2 = 2(m + 1).$$

■

2.2.6 More examples of proofs

Recall Theorem 2.2.5, which states that for every *even* integer $n \geq 4$, there exists a 3-regular graph with n vertices. The following theorem explains why we stated this theorem for even values of n.

Theorem 2.2.13 *Let $n \geq 5$ be an odd integer. There is no 3-regular graph with n vertices.*

Proof. The proof is by contradiction. So we assume that there exists a graph $G = (V, E)$ with n vertices that is 3-regular. Let m be the number of edges in G. Since $deg(v) = 3$ for every vertex, we have

$$\sum_{v \in V} deg(v) = 3n.$$

On the other hand, by Theorem 2.2.3, we have

$$\sum_{v \in V} deg(v) = 2m.$$

It follows that $3n = 2m$, which can be rewritten as $m = 3n/2$. Since m is an integer, and since $gcd(2, 3) = 1$, $n/2$ must be an integer. Hence, n is even, which is a contradiction. ∎

Let K_n be the *complete graph* on n vertices. This graph has a vertex set of size n, and every pair of distinct vertices is joined by an edge.

If $G = (V, E)$ is a graph with n vertices, then the *complement* $\overline{G}$ of G is the graph with vertex set V that consists of those edges of K_n that are not present in G.

Theorem 2.2.14 *Let $n \geq 2$ and let G be a graph on n vertices. Then G is connected or $\overline{G}$ is connected.*

Proof. We prove the theorem by induction on the number n of vertices. For the basis, assume that $n = 2$. There are two possibilities for the graph G:

1. G contains one edge. In this case, G is connected.

2. G does not contain an edge. In this case, the complement $\overline{G}$ contains one edge and, therefore, $\overline{G}$ is connected.

Thus, for $n = 2$, the theorem is true.

Let $n \geq 2$ and assume that the theorem is true for every graph with n vertices. Let G be graph with $n + 1$ vertices. We have to prove that G is connected or $\overline{G}$ is connected. We consider three cases.

Case 1: There is a vertex v whose degree in G is equal to n.

Since G has $n+1$ vertices, v is connected by an edge to every other vertex of G. Therefore, G is connected.

Case 2: There is a vertex v whose degree in G is equal to 0.

In this case, the degree of v in the graph $\overline{G}$ is equal to n. Since $\overline{G}$ has $n+1$ vertices, v is connected by an edge to every other vertex of $\overline{G}$. Therefore, $\overline{G}$ is connected.

Case 3: For every vertex v, the degree of v in G is in $\{1, 2, \ldots, n - 1\}$.

Let v be an arbitrary vertex of G. Let G' be the graph obtained by deleting from G the vertex v, together with all edges that are incident on v. Since G' has n vertices, we know from the induction hypothesis that G' is connected or $\overline{G'}$ is connected.

Let us first assume that G' is connected. Then the graph G is connected as well, because there is at least one edge in G between v and some vertex of G'.

If G' is not connected, then $\overline{G'}$ must be connected. Since we are in Case 3, we know that the degree of v in G is in the set $\{1, 2, \ldots, n-1\}$. It follows that the degree of v in the graph $\overline{G}$ is in this set as well. Hence, there is at least one edge in $\overline{G}$ between v and some vertex in $\overline{G'}$. This implies that $\overline{G}$ is connected. ∎

The previous theorem can be rephrased as follows:

Theorem 2.2.15 *Let $n \geq 2$ and consider the complete graph K_n on n vertices. Color each edge of this graph as either red or blue. Let R be the graph consisting of all the red edges, and let B be the graph consisting of all the blue edges. Then R is connected or B is connected.*

If you like surprising proofs of various mathematical results, you should read the book *Proofs from THE BOOK* by Martin Aigner and Günter Ziegler.

2.3 Asymptotic Notation

Let $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ be functions such that $f(n) > 0$ and $g(n) > 0$ for all $n \in \mathbb{N}$.

- We say that $f(n) = O(g(n))$ if the following is true: There exist constants $c > 0$ and $k > 0$ such that for all $n \geq k$,

$$f(n) \leq c \cdot g(n).$$

- We say that $f(n) = \Omega(g(n))$ if the following is true: There exist constants $c > 0$ and $k > 0$ such that for all $n \geq k$,

$$f(n) \geq c \cdot g(n).$$

- We say that $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. Thus, there exist constants $c > 0$, $c' > 0$, and $k > 0$ such that for all $n \geq k$,

$$c \cdot g(n) \leq f(n) \leq c' \cdot g(n).$$

For example, for all $n \geq 1$, we have

$$\begin{aligned} 13 + 7n - 5n^2 + 8n^3 &\leq 13 + 7n + 8n^3 \\ &\leq 13n^3 + 7n^3 + 8n^3 \\ &= 28n^3. \end{aligned}$$

Thus, by taking $c = 28$ and $k = 1$, it follows that

$$13 + 7n - 5n^2 + 8n^3 = O(n^3). \tag{2.1}$$

We also have

$$13 + 7n - 5n^2 + 8n^3 \geq -5n^2 + 8n^3.$$

Since $n^3 \geq 5n^2$ for all $n \geq 5$, it follows that, again for all $n \geq 5$,

$$\begin{aligned} 13 + 7n - 5n^2 + 8n^3 &\geq -5n^2 + 8n^3 \\ &\geq -n^3 + 8n^3 \\ &= 7n^3. \end{aligned}$$

Hence, by taking $c = 7$ and $k = 5$, we have shown that

$$13 + 7n - 5n^2 + 8n^3 = \Omega(n^3). \tag{2.2}$$

It follows from (2.1) and (2.2) that

$$13 + 7n - 5n^2 + 8n^3 = \Theta(n^3).$$

2.4 Logarithms

If b and x are positive real numbers, then $\log_b x$ is the logarithm of x with base b. Note that

$$\log_b x = y \text{ if and only if } b^y = x.$$

If $b = 2$, then we write $\log x$ instead of $\log_2 x$. We write $\ln x$ to refer to the natural logarithm of x with base e.

Lemma 2.4.1 *If $b > 0$ and $x > 0$, then*

$$b^{\log_b x} = x.$$

Proof. We have seen above that $y = \log_b x$ if and only if $b^y = x$. Thus, if we write $y = \log_b x$, then $b^{\log_b x} = b^y = x$. ∎

For example, if $x > 0$, then

$$2^{\log x} = x.$$

Lemma 2.4.2 *If $b > 0$, $x > 0$, and a is a real number, then*

$$\log_b(x^a) = a \log_b x.$$

Proof. Define $y = \log_b x$. Then

$$a \log_b x = ay.$$

Since $y = \log_b x$, we have $b^y = x$ and, thus,

$$x^a = (b^y)^a = b^{ay}.$$

Taking logarithms (with base b) on both sides gives

$$\log_b (x^a) = \log_b (b^{ay}) = ay = a \log_b x.$$

∎

For example, for $x > 1$, we get

$$2 \log \log x = \log \left(\log^2 x\right)$$

and

$$2^{2 \log \log x} = 2^{\log\left(\log^2 x\right)} = \log^2 x.$$

Lemma 2.4.3 *If $b > 0$, $c > 0$, and $x > 0$, then*

$$\log_b x = \frac{\log_c x}{\log_c b}.$$

Proof. Define $y = \log_b x$. Then $b^y = x$, and we get

$$\log_c x = \log_c (b^y) = y \log_c b = \log_b x \log_c b.$$

∎

For example, if $x > 0$, then

$$\log x = \frac{\ln x}{\ln 2}.$$

2.5 Exercises

2.1 Prove that $\sqrt{p}$ is irrational for every prime number p.

2.2 Let n be a positive integer that is not a perfect square. Prove that $\sqrt{n}$ is irrational.

2.3 Use induction to prove that every positive integer can be written as a product of prime numbers.

2.4 Prove by induction that $n^4 - 4n^2$ is divisible by 3, for all integers $n \geq 1$.

2.5 Prove that

$$\sum_{i=1}^{n} \frac{1}{i^2} < 2 - 1/n,$$

for every integer $n \geq 2$.

2.6 Prove that 9 divides $n^3 + (n+1)^3 + (n+2)^3$, for every integer $n \geq 0$.

2.7 The Fermat numbers $F_0, F_1, F_2, \ldots$ are defined by $F_n = 2^{2^n} + 1$ for $n \geq 0$.

- Prove by induction that

  $$F_0 F_1 F_2 \cdots F_{n-1} = F_n - 2$$

  for $n \geq 1$.

- Prove that for any two distinct integers $n \geq 0$ and $m \geq 0$, the greatest common divisor of F_n and F_m is equal to 1.

- Conclude that there are infinitely many prime numbers.

2.8 Prove by induction that $n! > 2^{1+n/2}$ for all integers $n \geq 3$.

Chapter 3

Counting

There are three types of people, those who can count and those who cannot count.

Given a set with 23 elements, how many subsets of size 17 are there? How many solutions are there to the equation

$$x_1 + x_2 + \cdots + x_{12} = 873,$$

where $x_1 \geq 0$, $x_2 \geq 0$, $\ldots$, $x_{12} \geq 0$ are integers? In this chapter, we will introduce some general techniques that can be used to answer questions of these types.

3.1 The Product Rule

How many strings of two characters are there, if the first character must be an uppercase letter and the second character must be a digit? Examples of such strings are $A0$, $K7$, and $Z9$. The answer is obviously $26 \cdot 10 = 260$, because there are 26 choices for the first character and, no matter which letter we choose for being the first character, there are 10 choices for the second character. We can look at this in the following way: Consider the "procedure" of writing down a string of two characters, the first one being an uppercase letter, and the second one being a digit. Then our original question becomes "how many ways are there to perform this procedure?" Observe that the procedure consists of two "tasks", the first one being writing

down the first character, and the second one being writing down the second character. Obviously, there are 26 ways to do the first task. Next, observe that, regardless of how we do the first task, there are 10 ways to do the second task. The Product Rule states that the total number of ways to perform the entire procedure is $26 \cdot 10 = 260$.

> **Product Rule:** Assume a procedure consists of performing a sequence of m tasks in order. Furthermore, assume that for each $i = 1, 2, \ldots, m$, there are N_i ways to do the i-th task, regardless of how the first $i - 1$ tasks were done. Then, there are $N_1 N_2 \cdots N_m$ ways to do the entire procedure.

In the example above, we have $m = 2$, $N_1 = 26$, and $N_2 = 10$.

3.1.1 Counting Bitstrings of Length n

Let $n \geq 1$ be an integer. A *bitstring* of length n is a sequence of 0's and 1's. How many bitstrings of length n are there? To apply the Product Rule, we have to specify the "procedure" and the "tasks":

- The procedure is "write down a bitstring of length n".
- For $i = 1, 2, \ldots, n$, the i-th task is "write down one bit".

There are two ways to do the i-th task, regardless of how we did the first $i - 1$ tasks. Therefore, we can apply the Product Rule with $N_i = 2$ for $i = 1, 2, \ldots, n$, and conclude that there are $N_1 N_2 \cdots N_n = 2^n$ ways to do the entire procedure. As a result, the number of bitstrings of length n is equal to 2^n.

Theorem 3.1.1 *For any integer $n \geq 1$, the number of bitstrings of length n is equal to 2^n.*

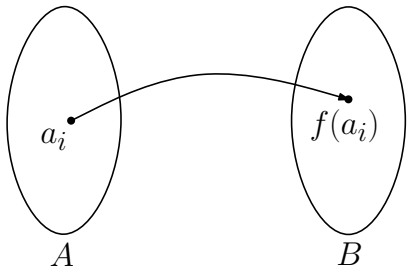
3.1.2 Counting Functions

Let $m \geq 1$ and $n \geq 1$ be integers, let A be a set of size m, and let B be a set of size n. How many functions $f : A \to B$ are there?

Write the set A as $A = \{a_1, a_2, \ldots, a_m\}$. Any function $f : A \to B$ is completely specified by the values $f(a_1), f(a_2), \ldots, f(a_m)$, where each such

value can be any element of B. Again, we are going to apply the Product Rule. Thus, we have to specify the "procedure" and the "tasks":

- The procedure is "specify the values $f(a_1), f(a_2), \ldots, f(a_m)$".
- For $i = 1, 2, \ldots, m$, the i-th task is "specify the value $f(a_i)$".



For each i, $f(a_i)$ can be any of the n elements of B. As a result, there are $N_i = n$ ways to do the i-th task, regardless of how we did the first $i-1$ tasks. By the Product Rule, there are $N_1 N_2 \cdots N_m = n^m$ ways to do the entire procedure and, hence, this many functions $f : A \to B$. We have proved the following result:

Theorem 3.1.2 *Let $m \geq 1$ and $n \geq 1$ be integers, let A be a set of size m, and let B be a set of size n. The number of functions $f : A \to B$ is equal to n^m.*

Recall that a function $f : A \to B$ is *one-to-one* if for any i and j with $i \neq j$, we have $f(a_i) \neq f(a_j)$. How many one-to-one functions $f : A \to B$ are there?

If $m > n$, then there is no such function. (Do you see why?) Assume that $m \leq n$. To determine the number of one-to-one functions, we use the same procedure and tasks as before.

- Since $f(a_1)$ can be any of the n elements of B, there are $N_1 = n$ ways to do the first task.
- In the second task, we have to specify the value $f(a_2)$. Since the function f is one-to-one and since we have already specified $f(a_1)$, we can choose $f(a_2)$ to be any of the $n-1$ elements in the set $B \setminus \{f(a_1)\}$. As a result, there are $N_2 = n - 1$ ways to do the second task. Note that this is true, regardless of how we did the first task.

- In general, in the i-th task, we have to specify the value $f(a_i)$. Since we have already specified $f(a_1), f(a_2), \ldots, f(a_{i-1})$, we can choose $f(a_i)$ to be any of the $n - i + 1$ elements in the set

$$B \setminus \{f(a_1), f(a_2), \ldots, f(a_{i-1})\}.$$

  As a result, there are $N_i = n - i + 1$ ways to do the i-th task. Note that this is true, regardless of how we did the first $i - 1$ tasks.

By the Product Rule, there are

$$N_1 N_2 \cdots N_m = n(n-1)(n-2)\cdots(n-m+1)$$

ways to do the entire procedure, which is also the number of one-to-one functions $f : A \to B$.

Recall the *factorial function*

$$k! = \begin{cases} 1 & \text{if } k = 0, \\ 1 \cdot 2 \cdot 3 \cdots k & \text{if } k \geq 1. \end{cases}$$

We can simplify the product

$$n(n-1)(n-2)\cdots(n-m+1)$$

by observing that it is "almost" a factorial:

$$\begin{aligned} & n(n-1)(n-2)\cdots(n-m+1) \\ &= n(n-1)(n-2)\cdots(n-m+1) \cdot \frac{(n-m)(n-m-1)\cdots 1}{(n-m)(n-m-1)\cdots 1} \\ &= \frac{n(n-1)(n-2)\cdots 1}{(n-m)(n-m-1)\cdots 1} \\ &= \frac{n!}{(n-m)!}. \end{aligned}$$

We have proved the following result:

Theorem 3.1.3 *Let $m \geq 1$ and $n \geq 1$ be integers, let A be a set of size m, and let B be a set of size n.*

1. *If $m > n$, then there is no one-to-one function $f : A \to B$.*
2. *If $m \leq n$, then the number of one-to-one functions $f : A \to B$ is equal to*

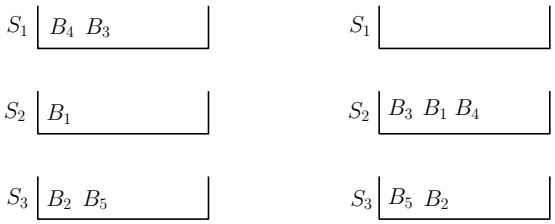
$$\frac{n!}{(n-m)!}.$$

3.1.3 Placing Books on Shelves

Let $m \geq 1$ and $n \geq 1$ be integers, and consider m books $B_1, B_2, \ldots, B_m$ and n book shelves $S_1, S_2, \ldots, S_n$. How many ways are there to place the books on the shelves? Placing the books on the shelves means that

- we specify for each book the shelf at which this book is placed, and
- we specify for each shelf the left-to-right order of the books that are placed on that shelf.

Some book shelves may be empty. We assume that each shelf is large enough to fit all books. In the figure below, you see two different placements.



We are again going to use the Product Rule to determine the number of placements.

- The procedure is "place the m books on the n shelves".
- For $i = 1, 2, \ldots, m$, the i-th task is "place book B_i on the shelves". When placing book B_i, we can place it on the far left or far right of any shelf or between any two of the books $B_1, \ldots, B_{i-1}$ that have already been placed.

Let us see how many ways there are to do each task.

- Just before we place book B_1, all shelves are empty. Therefore, there are $N_1 = n$ ways to do the first task.
- In the second task, we have to place book B_2. Since B_1 has already been placed, we have the following possibilities for placing B_2:
  - We place B_2 on the far left of any of the n shelves.
  - We place B_2 immediately to the right of B_1.

As a result, there are $N_2 = n + 1$ ways to do the second task. Note that this is true, regardless of how we did the first task.

- In general, in the i-th task, we have to place book B_i. Since $B_1, B_2, \ldots, B_{i-1}$ have already been placed, we have the following possibilities for placing B_i:
  - We place B_i on the far left of any of the n shelves.
  - We place B_i immediately to the right of one of $B_1, B_2, \ldots, B_{i-1}$.

  As a result, there are $N_i = n + i - 1$ ways to do the i-th task. Note that this is true, regardless of how we did the first $i - 1$ tasks.

Thus, by the Product Rule, there are

$$N_1 N_2 \cdots N_m = n(n+1)(n+2)\cdots(n+m-1)$$

ways to do the entire procedure, which is also the number of placements of the m books on the n shelves. As before, we use factorials to simplify this product:

$$\begin{aligned} & n(n+1)(n+2)\cdots(n+m-1) \\ = \; & \frac{1\cdot 2\cdot 3\cdots(n-1)}{1\cdot 2\cdot 3\cdots(n-1)} \cdot n(n+1)(n+2)\cdots(n+m-1) \\ = \; & \frac{(n+m-1)!}{(n-1)!}. \end{aligned}$$

We have proved the following result:

Theorem 3.1.4 *Let $m \geq 1$ and $n \geq 1$ be integers. The number of ways to place m books on n book shelves is equal to*

$$\frac{(n+m-1)!}{(n-1)!}.$$

3.2 The Bijection Rule

Let $n \geq 0$ be an integer and let S be a set with n elements. How many subsets does S have? If $n = 0$, then $S = \emptyset$ and there is only one subset of S, namely S itself. Assume from now on that $n \geq 1$. As we will see below,

asking for the number of subsets of S is exactly the same as asking for the number of bitstrings of length n.

Let A and B be finite sets. Recall that a function $f : A \to B$ is a *bijection* if

- f is one-to-one, i.e., if $a \neq a'$ then $f(a) \neq f(a')$, and
- f is onto, i.e., for each b in B, there is an a in A such that $f(a) = b$.

This means that

- each element of A corresponds to a unique element of B and
- each element of B corresponds to a unique element of A.

It should be clear that this means that A and B contain the same number of elements.

Bijection Rule: Let A and B be finite sets. If there exists a bijection $f : A \to B$, then $|A| = |B|$, i.e., A and B have the same size.

Let us see how we can apply this rule to the subset problem. We define the following two sets A and B:

- $A = \mathcal{P}(S)$, i.e., the power set of S, which is the set of all subsets of S:

$$\mathcal{P}(S) = \{T : T \subseteq S\}.$$

- B is the set of all bitstrings of length n.

We have seen in Theorem 3.1.1 that the set B has size 2^n. Therefore, if we can show that there exists a bijection $f : A \to B$, then, according to the Bijection Rule, we have $|A| = |B|$ and, thus, the number of subsets of S is equal to 2^n.

Write the set S as $S = \{s_1, s_2, \ldots, s_n\}$. We define the function $f : A \to B$ in the following way:

- For any $T \in A$ (i.e., $T \subseteq S$), $f(T)$ is the bitstring $b_1 b_2 \ldots b_n$, where

$$b_i = \begin{cases} 1 & \text{if } s_i \in T, \\ 0 & \text{if } s_i \notin T. \end{cases}$$

For example, assume that $n = 5$.

- If $T = \{s_1, s_3, s_4\}$, then $f(T) = 10110$.
- If $T = S = \{s_1, s_2, s_3, s_4, s_5\}$, then $f(T) = 11111$.
- If $T = \emptyset$, then $f(T) = 00000$.

It is not difficult to see that each subset of S corresponds to a unique bitstring of length n, and each bitstring of length n corresponds to a unique subset of S. Therefore, this function f is a bijection between A and B.

Thus, we have shown that there exists a bijection $f : A \to B$. This, together with Theorem 3.1.1 and the Bijection Rule, implies the following result:

Theorem 3.2.1 *For any integer $n \geq 0$, the number of subsets of a set of size n is equal to 2^n.*

You will probably have noticed that we could have proved this result directly using the Product Rule: The procedure "specify a subset of $S = \{s_1, s_2, \ldots, s_n\}$" can be carried out by specifying, for $i = 1, 2, \ldots, n$, whether or not s_i is contained in the subset. For each i, there are two choices. As a result, there are 2^n ways to do the procedure.

3.3 The Complement Rule

Consider strings consisting of 8 characters, each character being a lowercase letter or a digit. Such a string is called a *valid password* if it contains at least one digit. How many valid passwords are there? One way to answer this question is to first count the valid passwords with exactly one digit, then count the valid passwords with exactly two digits, then count the valid passwords with exactly three digits, etc. As we will see below, it is easier to first count the strings that do *not* contain any digit.

Recall that the *difference* $U \setminus A$ of the two sets U and A is defined as

$$U \setminus A = \{x : x \in U \text{ and } x \notin A\}.$$

Complement Rule: Let U be a finite set and let A be a subset of U. Then

$$|A| = |U| - |U \setminus A|.$$

This rule follows easily from the fact that $|U| = |A| + |U \setminus A|$, which holds because each element in U is either in A or in $U \setminus A$.

To apply the Complement Rule to the password problem, let U be the set of all strings consisting of 8 characters, each character being a lowercase letter or a digit, and let A be the set of all valid passwords, i.e., all strings in U that contain at least one digit. Note that $U \setminus A$ is the set of all strings of 8 characters, each character being a lowercase letter or a digit, that do not contain any digit. In other words, $U \setminus A$ is the set of all strings of 8 characters, where each character is a lowercase letter.

By the Product Rule, the set U has size 36^8, because each string in U has 8 characters, and there are $26 + 10 = 36$ choices for each character. Similarly, the set $U \setminus A$ has size 26^8, because there are 26 choices for each of the 8 characters. Then, by the Complement Rule, the number of valid passwords is equal to

$$|A| = |U| - |U \setminus A| = 36^8 - 26^8 = 2,612,282,842,880.$$

3.4 The Sum Rule

If A and B are two finite sets that are *disjoint*, i.e., $A \cap B = \emptyset$, then it is obvious that the size of $A \cup B$ is equal to the sum of the sizes of A and B.

Sum Rule: Let $A_1, A_2, \ldots, A_m$ be a sequence of finite and pairwise disjoint sets. Then

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

Note that we already used this rule in Section 3.3 when we argued why the Complement Rule is correct!

To give an example, consider strings consisting of 6, 7, or 8 characters, each character being a lowercase letter or a digit. Such a string is called a *valid password* if it contains at least one digit. Let A be the set of all valid passwords. What is the size of A?

For $i = 6, 7, 8$, let A_i be the set of all valid passwords of length i. It is obvious that $A = A_6 \cup A_7 \cup A_8$. Since the three sets A_6, A_7, and A_8 are pairwise disjoint, we have, by the Sum Rule,

$$|A| = |A_6| + |A_7| + |A_8|.$$

We have seen in Section 3.3 that $|A_8| = 36^8 - 26^8$. By the same arguments, we have $|A_6| = 36^6 - 26^6$ and $|A_7| = 36^7 - 26^7$. Thus, the number of valid passwords is equal to

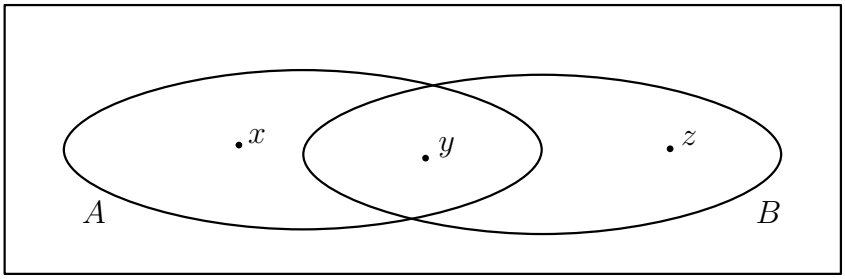
$$|A| = \left(36^6 - 26^6\right) + \left(36^7 - 26^7\right) + \left(36^8 - 26^8\right) = 2,684,483,063,360.$$

3.5 The Principle of Inclusion and Exclusion

The Sum Rule holds only for sets that are pairwise disjoint. Consider two finite sets A and B that are not necessarily disjoint. How can we determine the size of the union $A \cup B$? We can start with the sum $|A| + |B|$, i.e., we *include* both A and B. In the figure below,

- x is in A but not in B; it is counted exactly once in $|A| + |B|$,
- z is in B but not in A; it is counted exactly once in $|A| + |B|$,
- y is in A and in B; it is counted exactly twice in $|A| + |B|$.

Based on this, if we subtract the size of the intersection $A \cap B$, i.e., we *exclude* $A \cap B$, then we have counted every element of $A \cup B$ exactly once.



Inclusion-Exclusion: Let A and B be finite sets. Then

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

To give an example, let us count the bitstrings of length 17 that start with 010 or end with 11. Let S be the set of all such bitstrings. Define A to be the set of all bitstrings of length 17 that start with 010, and define B to be the set of all bitstrings of length 17 that end with 11. Then $S = A \cup B$ and, thus, we have to determine the size of $A \cup B$.

- The size of A is equal to the number of bitstrings of length 14, because the first three bits of every string in A are fixed. Therefore, by the Product Rule, we have $|A| = 2^{14}$.
- The size of B is equal to the number of bitstrings of length 15, because the last two bits of every string in B are fixed. Therefore, by the Product Rule, we have $|B| = 2^{15}$.
- Each string in $A \cap B$ starts with 010 *and* ends with 11. Thus, five bits are fixed for every string in $A \cap B$. It follows that the size of $A \cap B$ is equal to the number of bitstrings of length 12. Therefore, by the Product Rule, we have $|A \cap B| = 2^{12}$.

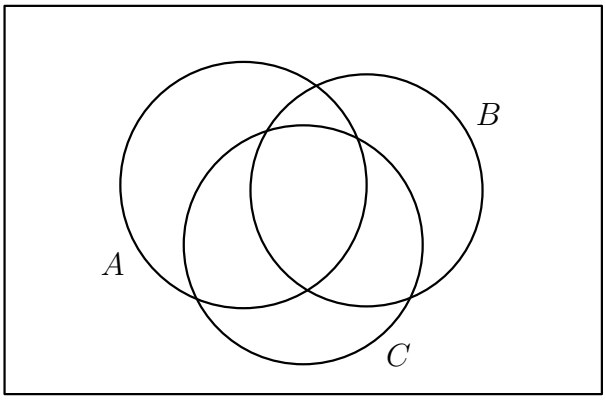
By applying the Inclusion-Exclusion formula, it follows that

$$|S| = |A \cup B| = |A| + |B| - |A \cap B| = 2^{14} + 2^{15} - 2^{12} = 45,056.$$

The Inclusion-Exclusion formula can be generalized to more than two sets. You are encouraged to verify, using the figure below, that the following formula is the correct one for three sets.

Inclusion-Exclusion: Let A, B, and C be finite sets. Then

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$



To give an example, how many bitstrings of length 17 are there that start with 010, or end with 11, or have 10 at positions 7 and 8? Let S be the set of all such bitstrings. Define A to be the set of all bitstrings of length 17 that

start with 010, define B to be the set of all bitstrings of length 17 that end with 11, and define C to be the set of all bitstrings of length 17 that have 10 at positions 7 and 8. Then $S = A \cup B \cup C$ and, thus, we have to determine the size of $A \cup B \cup C$.

- We have seen before that $|A| = 2^{14}$, $|B| = 2^{15}$, and $|A \cap B| = 2^{12}$.
- We have $|C| = 2^{15}$, because the bits at positions 7 and 8 are fixed for every string in C.
- We have $|A \cap C| = 2^{12}$, because 5 bits are fixed for every string in $A \cap C$.
- We have $|B \cap C| = 2^{13}$, because 4 bits are fixed for every string in $B \cap C$.
- We have $|A \cap B \cap C| = 2^{10}$, because 7 bits are fixed for every string in $A \cap B \cap C$.

By applying the Inclusion-Exclusion formula, it follows that

$$\begin{aligned} |S| &= |A \cup B \cup C| \\ &= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| \\ &= 2^{14} + 2^{15} + 2^{15} - 2^{12} - 2^{12} - 2^{13} + 2^{10} \\ &= 66,560. \end{aligned}$$

3.6 Permutations and Binomial Coefficients

A *permutation* of a finite set S is an *ordered* sequence of the elements of S, in which each element occurs exactly once. For example, the set $S = \{a, b, c\}$ has exactly six permutations:

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

Theorem 3.6.1 *Let $n \geq 0$ be an integer and let S be a set with n elements. There are exactly $n!$ many permutations of S.*

Proof. If $n = 0$, then $S = \emptyset$ and the only permutation of S is the empty sequence. Since $0! = 1$, the claim holds for $n = 0$. Assume that $n \geq 1$ and denote the elements of S by $s_1, s_2, \ldots, s_n$. Consider the procedure "write

down a permutation of S" and, for $i = 1, 2, \ldots, n$, the task "write down the i-th element in the permutation". When we do the i-th task, we have already written down $i-1$ elements of the permutation; we cannot take any of these elements for the i-th task. Therefore, there are $n-(i-1) = n-i+1$ ways to do the i-th task. By the Product Rule, there are

$$n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n!$$

ways to do the procedure. This number is equal to the number of permutations of S. ∎

Note that we could also have used Theorem 3.1.3 to prove Theorem 3.6.1: A permutation of S is a one-to-one function $f : S \to S$. Therefore, by applying Theorem 3.1.3 with $A = S$, $B = S$ and, thus, $m = n$, we obtain Theorem 3.6.1.

Consider the set $S = \{a, b, c, d, e\}$. How many 3-element subsets does S have? Recall that in a set, the order of the elements does not matter. Here is a list of all 10 subsets of S having size 3:

$$\{a, b, c\}, \{a, b, d\}, \{a, b, e\}, \{a, c, d\}, \{a, c, e\},$$
$$\{a, d, e\}, \{b, c, d\}, \{b, c, e\}, \{b, d, e\}, \{c, d, e\}$$

Definition 3.6.2 Let n and k be integers. We define the *binomial coefficient* $\binom{n}{k}$ to be the number of k-element subsets of an n-element set.

The symbol $\binom{n}{k}$ is pronounced as "n choose k".

The example above shows that $\binom{5}{3} = 10$. Since the empty set has exactly one subset of size zero (the empty set itself), we have $\binom{0}{0} = 1$. Note that $\binom{n}{k} = 0$ if $k < 0$ or $k > n$. Below, we derive a formula for the value of $\binom{n}{k}$ if $0 \leq k \leq n$.

Let S be a set with n elements and let A be the set of all *ordered* sequences consisting of exactly k pairwise distinct elements of S. We are going to count the number of elements of A in two different ways.

The first way is by using the Product Rule. This gives

$$|A| = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}. \tag{3.1}$$

In the second way, we do the following:

- Write down all $\binom{n}{k}$ subsets of S having size k.
- For each of these subsets, write down a list of all $k!$ permutations of this subset.

If we put all these lists together, then we obtain a big list in which each ordered sequence of k pairwise distinct elements of S appears exactly once. In other words, the big list contains each element of A exactly once. Since the big list has size $\binom{n}{k}k!$, it follows that

$$|A| = \binom{n}{k}k! \tag{3.2}$$

By equating the right-hand sides of (3.1) and (3.2), we obtain the following result:

Theorem 3.6.3 *Let n and k be integers with $0 \leq k \leq n$. Then*

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

For example,

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2} = 10$$

and

$$\binom{0}{0} = \frac{0!}{0!0!} = \frac{1}{1 \cdot 1} = 1;$$

recall that we defined 0! to be equal to 1.

3.6.1 Some Examples

Consider a standard deck of 52 cards. How many hands of 5 cards are there? Any such hand is a 5-element subset of the set of 52 cards and, therefore, the number of hands of 5 cards is equal to

$$\binom{52}{5} = \frac{52!}{5!47!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2,598,960.$$

Let n and k be integers with $0 \leq k \leq n$. How many bitstrings of length n have exactly k many 1s? We can answer this question using the Product Rule:

- The procedure is "write down a bitstring of length n having exactly k many 1s".
- Task 1: Choose k positions out of $1, 2, \ldots, n$.
- Task 2: Write a 1 in each of the k chosen positions.
- Task 3: Write a 0 in each of the $n - k$ remaining positions.

There are $\binom{n}{k}$ ways to do the first task, there is only one way to do the second task, and there is only one way to do the third task. Thus, by the Product Rule, the number of ways to do the procedure and, therefore, the number of bitstrings of length n having exactly k many 1s, is equal to

$$\binom{n}{k} \cdot 1 \cdot 1 = \binom{n}{k}.$$

We can also use the Bijection Rule, by observing that there is bijection between

- the set of all bitstrings of length n having exactly k many 1s, and
- the set of all k-element subsets of an n-element set.

Since the latter set has size $\binom{n}{k}$, the former set has size $\binom{n}{k}$ as well.

Theorem 3.6.4 *Let n and k be integers with $0 \leq k \leq n$. The number of bitstrings of length n having exactly k many 1s is equal to $\binom{n}{k}$.*

3.6.2 Newton's Binomial Theorem

Consider

$$(x + y)^5 = (x + y)(x + y)(x + y)(x + y)(x + y).$$

If we expand the expression on the right-hand side, we get terms

$$x^5, x^4y, x^3y^2, x^2y^3, xy^4, y^5,$$

each with some coefficient. What is the coefficient of x^2y^3? We obtain a term x^2y^3, by

- choosing 3 of the 5 terms $x + y$,

- taking y in each of the 3 chosen terms $x+y$, and
- taking x in each of the other 2 terms $x+y$.

Since there are $\binom{5}{3}$ ways to do this, the coefficient of x^2y^3 is equal to $\binom{5}{3} = 10$.

Theorem 3.6.5 (Newton's Binomial Theorem) *For any integer $n \geq 0$, we have*

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

Proof. The expression $(x+y)^n$ is a product of n terms $x+y$. By expanding this product, we get a term $x^{n-k}y^k$ for each $k = 0, 1, \ldots, n$. We get a term $x^{n-k}y^k$ by

- choosing k of the n terms $x+y$,
- taking y in each of the k chosen terms $x+y$, and
- taking x in each of the other $n-k$ terms $x+y$.

Since there are $\binom{n}{k}$ ways to do this, the coefficient of $x^{n-k}y^k$ is equal to $\binom{n}{k}$. ∎

For example, we have

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

To determine the coefficient of $x^{12}y^{13}$ in $(x+y)^{25}$, we take $n = 25$ and $k = 13$ in Newton's Binomial Theorem, and get $\binom{25}{13}$.

What is the coefficient of $x^{12}y^{13}$ in $(2x-5y)^{25}$? Observe that

$$(2x-5y)^{25} = ((2x) + (-5y))^{25}.$$

By replacing x by $2x$, and y by $-5y$ in Newton's Binomial Theorem, we get

$$(2x-5y)^{25} = \sum_{k=0}^{25} \binom{25}{k} (2x)^{25-k}(-5y)^k.$$

By taking $k = 13$, we obtain the coefficient of $x^{12}y^{13}$:

$$\binom{25}{13} \cdot 2^{25-13} \cdot (-5)^{13} = -\binom{25}{13} \cdot 2^{12} \cdot 5^{13}.$$

Newton's Binomial Theorem leads to identities for summations involving binomial coefficients:

Theorem 3.6.6 *For any integer $n \geq 0$, we have*

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

Proof. Take $x = y = 1$ in Newton's Binomial Theorem. ∎

In Section 3.7, we will see a proof of Theorem 3.6.6 that does not use Newton's Binomial Theorem.

Theorem 3.6.7 *For any integer $n \geq 1$, we have*

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

Proof. Take $x = 1$ and $y = -1$ in Newton's Binomial Theorem. ∎

3.7 Combinatorial Proofs

In a combinatorial proof, we show the validity of an identity, such as the one in Theorem 3.6.6, by interpreting it as a counting problem. In this section, we will give several examples.

Theorem 3.7.1 *For integers n and k with $0 \leq k \leq n$, we have*

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

Proof. The claim can be proved using Theorem 3.6.3. To obtain a combinatorial proof, let S be a set with n elements. Recall that

- $\binom{n}{k}$ is the number of ways to choose k elements from the set S,

which is the same as

- the number of ways to *not* choose $n-k$ elements from the set S.

The latter number is equal to $\binom{n}{n-k}$.

We can also prove the claim using Theorem 3.6.4:

- The number of bitstrings of length n with exactly k many 1s is equal to $\binom{n}{k}$.
- The number of bitstrings of length n with exactly $n-k$ many 0s is equal to $\binom{n}{n-k}$.

Since these two quantities are equal, the theorem follows. ∎

Theorem 3.7.2 (Pascal's Identity) *For integers n and k with $n \geq k \geq 1$, we have*

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

Proof. As in the previous theorem, the claim can be proved using Theorem 3.6.3. To obtain a combinatorial proof, let S be a set with $n+1$ elements. We are going to count the number of k-element subsets of S in two different ways.

First, by definition, the number of k-element subsets of S is equal to

$$\binom{n+1}{k}. \tag{3.3}$$

For the second way, we choose an element x in S and consider the set $T = S \setminus \{x\}$, i.e., the set obtained by removing x from S. Any k-element subset of S is of one of the following two types:

- The k-element subset of S does not contain x.
  - Any such subset is a k-element subset of T. Since T has size n, there are $\binom{n}{k}$ many k-element subsets of S that do not contain x.
- The k-element subset of S contains x.

- If A is any such subset, then $B = A \setminus \{x\}$ is a $(k-1)$-element subset of T.
- Conversely, for any $(k-1)$-element subset B of T, the set $A = B \cup \{x\}$ is a k-element subset of S that contains x.
- It follows that the number of k-element subsets of S containing x is equal to the number of $(k-1)$-element subsets of T. The latter number is equal to $\binom{n}{k-1}$.

Thus, the second way of counting shows that the number of k-element subsets of S is equal to

$$\binom{n}{k} + \binom{n}{k-1}. \tag{3.4}$$

Since the expressions in (3.3) and (3.4) count the same objects, they must be equal. Therefore, the proof is complete. ∎

Theorem 3.7.3 *For any integer $n \geq 0$, we have*

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

Proof. We have seen in Theorem 3.6.6 that this identity follows from Newton's Binomial Theorem. Below, we give a combinatorial proof.

Consider a set S with n elements. According to Theorem 3.2.1, this set has 2^n many subsets. A different way to count the subsets of S is by dividing them into groups according to their sizes. For each k with $0 \leq k \leq n$, consider all k-element subsets of S. The number of such subsets is equal to $\binom{n}{k}$. If we take the sum of all these binomial coefficients, then we have counted each subset of S exactly once. Thus,

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

is equal to the total number of subsets of S. ∎

Theorem 3.7.4 (Vandermonde's Identity) *For integers $m \geq 0$, $n \geq 0$, and $r \geq 0$ with $r \leq m$ and $r \leq n$, we have*

$$\sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}.$$

Proof. Consider a set S with $m+n$ elements. We are going to count the r-element subsets of S in two different ways.

First, by using the definition of binomial coefficients, the number of r-element subsets of S is equal to $\binom{m+n}{r}$.

For the second way, we partition the set S into two subsets A and B, where A has size m and B has size n. Observe that any r-element subset of S contains

- some elements, say k, of A, and
- $r-k$ elements of B.

Let k be any integer with $0 \leq k \leq r$, and let N_k be the number of r-element subsets of S that contain exactly k elements of A (and, thus, $r-k$ elements of B). Then,

$$\sum_{k=0}^{r} N_k = \binom{m+n}{r}.$$

To determine N_k, we use the Product Rule: We obtain any subset that is counted in N_k, by

- choosing k elements in A (there are $\binom{m}{k}$ ways to do this) and
- choosing $r-k$ elements in B (there are $\binom{n}{r-k}$ ways to do this).

It follows that

$$N_k = \binom{m}{k}\binom{n}{r-k}.$$

∎

Corollary 3.7.5 *For any integer $n \geq 0$, we have*

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

Proof. By taking $m = n = r$ in Vandermonde's Identity, we get

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

Using Theorem 3.7.1, we get

$$\binom{n}{k}\binom{n}{n-k} = \binom{n}{k}\binom{n}{k} = \binom{n}{k}^2.$$

■

3.8 Pascal's Triangle

> The computational method at the heart of Pascal's work was actually discovered by a Chinese mathematician named Jia Xian around 1050, published by another Chinese mathematician, Zhu Shijie, in 1303, discussed in a work by Cardano in 1570, and plugged into the greater whole of probability theory by Pascal, who ended up getting most of the credit.
> — Leonard Mlodinow, *The Drunkard's Walk*, 2008

We have seen that

- $\binom{n}{0} = 1$ for all integers $n \geq 0$,
- $\binom{n}{n} = 1$ for all integers $n \geq 0$,
- $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ for all integers $n \geq 2$ and k with $1 \leq k \leq n-1$; see Theorem 3.7.2.

These relations lead to an algorithm for generating binomial coefficients:

Algorithm GENERATEBINOMCOEFF**:**

BCoeff$(0,0) = 1$;
for $n = 1, 2, 3, \ldots$
do *BCoeff*$(n,0) = 1$;
  for $k = 1$ **to** $n-1$
  do *BCoeff*$(n,k) =$ *BCoeff*$(n-1,k-1) +$ *BCoeff*$(n-1,k)$
  endfor;
  BCoeff$(n,n) = 1$
endfor

The values $BCoeff(n,k)$ that are computed by this (non-terminating) algorithm satisfy

$$BCoeff(n,k) = \binom{n}{k} \text{ for } 0 \leq k \leq n.$$

The triangle obtained by arranging these binomial coefficients, with the n-th row containing all values $\binom{n}{0}$, $\binom{n}{1}$, ..., $\binom{n}{n}$, is called *Pascal's Triangle*. The figure below shows rows $0, 1, \ldots, 6$:

$$\binom{0}{0}$$

$$\binom{1}{0} \quad \binom{1}{1}$$

$$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$$

$$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$$

$$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$$

$$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$$

$$\binom{6}{0} \quad \binom{6}{1} \quad \binom{6}{2} \quad \binom{6}{3} \quad \binom{6}{4} \quad \binom{6}{5} \quad \binom{6}{6}$$

We obtain the values for the binomial coefficients by using the following rules:

- Each value along the boundary is equal to 1.
- Each value in the interior is equal to the sum of the two values above it.

In Figure 3.1, you see rows $0, 1, \ldots, 12$.

Below, we state some of our earlier results using Pascal's Triangle.

- The values in the n-th row are equal to the coefficients in Newton's Binomial Theorem. For example, the coefficients in the expansion of $(x+y)^5$ are given in the 5-th row:

$$(x+y)^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

1 10 45 120 210 252 210 120 45 10 1

1 11 55 165 330 462 462 330 165 55 11 1

1 12 66 220 495 792 924 792 495 220 66 12 1

Figure 3.1: Rows $0, 1, \ldots, 12$ of Pascal's Triangle.

- Theorem 3.6.6 states that the sum of all values in the n-th row is equal to 2^n.
- Theorem 3.7.1 states that reading the n-th row from left to right gives the same sequence as reading this row from right to left.
- Corollary 3.7.5 states that the sum of the squares of all values in the n-th row is equal to the middle element in the $2n$-th row.

3.9 More Counting Problems

3.9.1 Reordering the Letters of a Word

How many different strings can be made by reordering the letters of the 7-letter word

SUCCESS.

It should be clear that the answer is not 7!: If we swap, for example, the two occurrences of C, then we obtain the same string.

The correct answer can be obtained by applying the Product Rule. We start by counting the frequencies of each letter:

- The letter S occurs 3 times.
- The letter C occurs 2 times.
- The letter U occurs 1 time.
- The letter E occurs 1 time.

To apply the Product Rule, we have to specify the procedure and the tasks:

- The procedure is "write down the letters occurring in the word SUCCESS".
- The first task is "choose 3 positions out of 7, and write the letter S in each chosen position".
- The second task is "choose 2 positions out of the remaining 4, and write the letter C in each chosen position".

- The third task is "choose 1 position out of the remaining 2, and write the letter U in the chosen position".
- The fourth task is "choose 1 position out of the remaining 1, and write the letter E in the chosen position".

Since there are $\binom{7}{3}$ ways to do the first task, $\binom{4}{2}$ ways to do the second task, $\binom{2}{1}$ ways to do the third task, and $\binom{1}{1}$ way to do the fourth task, it follows that the total number of different strings that can be made by reordering the letters of the word SUCCESS is equal to

$$\binom{7}{3}\binom{4}{2}\binom{2}{1}\binom{1}{1} = 420.$$

In the four tasks above, we first chose the positions for the letter S, then the positions for the letter C, then the position for the letter U, and finally the position for the letter E. If we change the order, then we obtain the same answer. For example, if we choose the positions for the letters in the order C, E, U, S, then we obtain

$$\binom{7}{2}\binom{5}{1}\binom{4}{1}\binom{3}{3},$$

which is indeed equal to 420.

3.9.2 Counting Solutions of Linear Equations

Consider the equation

$$x_1 + x_2 + x_3 = 11.$$

We are interested in the number of solutions (x_1, x_2, x_3), where $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ are integers. Examples of solutions are

$$(2, 3, 6), (3, 2, 6), (0, 11, 0), (2, 0, 9).$$

Observe that we consider $(2, 3, 6)$ and $(3, 2, 6)$ to be different solutions.

We are going to use the Bijection Rule to determine the number of solutions. For this, we define A to be the set of all solutions, i.e.,

$$A = \{(x_1, x_2, x_3) : \ x_1 \geq 0, \, x_2 \geq 0, \, x_3 \geq 0 \text{ are integers}, \, x_1 + x_2 + x_3 = 11\}.$$

To apply the Bijection Rule, we need a set B and a bijection $f : A \to B$, such that it is easy to determine the size of B. This set B should be chosen such that its elements "encode" the elements of A in a unique way. Consider the following set B:

- B is the set of all binary strings of length 13 that contain exactly 2 many 1s (and, thus, exactly 11 many 0s).

The function $f : A \to B$ is defined as follows: If (x_1, x_2, x_3) is an element of A, then $f(x_1, x_2, x_3)$ is the binary string

- that starts with x_1 many 0s,
- is followed by one 1,
- is followed by x_2 many 0s,
- is followed by one 1,
- and ends with x_3 many 0s.

For example, we have

$$f(2, 3, 6) = 0010001000000,$$
$$f(3, 2, 6) = 0001001000000,$$
$$f(0, 11, 0) = 1000000000001,$$

and

$$f(2, 0, 9) = 0011000000000.$$

Observe that we have to verify that the string $f(x_1, x_2, x_3)$ belongs to the set B. This follows from the following observations:

- The string $f(x_1, x_2, x_3)$ contains exactly 2 many 1s.
- The number of 0s in the string $f(x_1, x_2, x_3)$ is equal to $x_1 + x_2 + x_3$, which is equal to 11, because (x_1, x_2, x_3) belongs to the set A.

It should be clear that this function f is one-to-one: If we take two different elements (x_1, x_2, x_3) in A, then f gives us two different bitstrings $f(x_1, x_2, x_3)$.

To prove that f is onto, we have to show that for every bitstring b in the set B, there is an element (x_1, x_2, x_3) in A such that $f(x_1, x_2, x_3) = b$. This element of A is obtained by taking

- x_1 to be the number of 0s to the left of the first 1,
- x_2 to be the number of 0s between the two 1s, and
- x_3 to be the number of 0s to the right of the second 1.

For example, if $b = 0000110000000$, then $x_1 = 4$, $x_2 = 0$, and $x_3 = 7$. Note that, since b has length 13 and contains exactly 2 many 1s, we have $x_1 + x_2 + x_3 = 11$ and, therefore, $(x_1, x_2, x_3) \in A$.

Thus, we have shown that $f : A \to B$ is indeed a bijection. We know from Theorem 3.6.4 that B has size $\binom{13}{2}$. Therefore, it follows from the Bijection Rule that

$$|A| = |B| = \binom{13}{2} = 78.$$

The following theorem states this result for general linear equations. You are encouraged to come up with the proof.

Theorem 3.9.1 *Let $k \geq 1$ and $n \geq 0$ be integers. The number of solutions to the equation*

$$x_1 + x_2 + \cdots + x_k = n,$$

where $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_k \geq 0$ are integers, is equal to

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

Let us now consider inequalities instead of equations. For example, consider the inequality

$$x_1 + x_2 + x_3 \leq 11.$$

Again, we are interested in the number of solutions (x_1, x_2, x_3), where $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ are integers. This inequality contains the same solutions as before, but it has additional solutions such as

$$(2, 3, 5), (3, 2, 5), (0, 1, 0), (0, 0, 0).$$

As before, we are going to apply the Bijection Rule. We define

$$A = \{(x_1, x_2, x_3) : \ x_1 \geq 0, \, x_2 \geq 0, \, x_3 \geq 0 \text{ are integers}, \, x_1 + x_2 + x_3 \leq 11\}$$

and B to be the set of all binary strings of length 14 that contain exactly 3 many 1s (and, thus, exactly 11 many 0s).

The function $f : A \to B$ is defined as follows: If (x_1, x_2, x_3) is an element of A, then $f(x_1, x_2, x_3)$ is the binary string

- that starts with x_1 many 0s,
- is followed by one 1,
- is followed by x_2 many 0s,
- is followed by one 1,
- is followed by x_3 many 0s,
- is followed by one 1,
- and ends with $14 - (x_1 + x_2 + x_3 + 3)$ many 0s.

For example, we have

$$f(2, 3, 6) = 00100010000001,$$

$$f(2, 3, 5) = 00100010000010,$$

$$f(0, 1, 0) = 10110000000000,$$

and

$$f(0, 0, 0) = 11100000000000.$$

As before, it can be verified that the string $f(x_1, x_2, x_3)$ belongs to the set B and the function f is a bijection. It then follows from the Bijection Rule that

$$|A| = |B| = \binom{14}{3} = 364.$$

The next theorem gives the answer for the general case. As before, you are encouraged to give a proof.

Theorem 3.9.2 *Let $k \geq 1$ and $n \geq 0$ be integers. The number of solutions to the inequality*

$$x_1 + x_2 + \cdots + x_k \leq n,$$

where $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_k \geq 0$ are integers, is equal to

$$\binom{n+k}{k}.$$

3.10 The Pigeonhole Principle

In any group of 366 people, there must be two people having the same birthday: Since there are 365 days in a year (ignoring leap years), it is not possible that the birthdays of 366 people are all distinct.

Pigeonhole Principle: Let $k \geq 1$ be an integer. If $k+1$ or more objects are placed into k boxes, then there is at least one box containing two or more objects.
Equivalently, if A and B are two finite sets such that $|A| > |B|$, then there is no one-to-one function from A to B.

3.10.1 India Pale Ale

Simon Pratt[1] loves to drink India Pale Ale (IPA)[2]. During each day of the month of April (which has 30 days), Simon drinks at least one bottle of IPA. Also, during this entire month, he drinks exactly 45 bottles of IPA. The claim is that there must be a sequence of consecutive days in April, during which Simon drinks exactly 14 bottles of IPA.

To prove this, let b_i be the number of bottles that Simon drinks on April i, for $i = 1, 2, \ldots, 30$. We are given that each b_i is a positive integer and

$$b_1 + b_2 + \cdots + b_{30} = 45.$$

Define, for $i = 1, 2, \ldots, 30$,

$$a_i = b_1 + b_2 + \cdots + b_i,$$

i.e., a_i is the total number of bottles of IPA that Simon drinks during the first i days of April. Consider the sequence of 60 numbers

$$a_1, a_2, \ldots, a_{30}, a_1 + 14, a_2 + 14, \ldots, a_{30} + 14.$$

Each number in this sequence is an integer that belongs to the set

$$\{1, 2, \ldots, 59\}.$$

[1]President of the Carleton Computer Science Society (2013–2014).

[2]Lindsay Bangs, President of the Carleton Computer Science Society (2014–2015), prefers wheat beer.

Therefore, by the Pigeonhole Principle, these 60 numbers cannot all be distinct. Observe that there are no duplicates in the sequence $a_1, a_2, \ldots, a_{30}$, because all b_i are at least one. Similarly, there are no duplicates in the sequence $a_1 + 14, a_2 + 14, \ldots, a_{30} + 14$. It follows that there are two indices i and j such that

$$a_i = a_j + 14.$$

Observe that j must be less than i and

$$14 = a_i - a_j = b_{j+1} + b_{j+2} + \cdots + b_i.$$

Thus, in the period from April $j + 1$ until April i, Simon drinks exactly 14 bottles of IPA.

3.10.2 Sequences Containing Divisible Numbers

Let $A = \{1, 2, \ldots, 2n\}$ and consider the sequence $n + 1, n + 2, \ldots, 2n$ of elements in A. This sequence has the property that none of its elements divides any other element in the sequence. Note that the sequence has length n. The following theorem states that such a sequence of length $n + 1$ does not exist.

Theorem 3.10.1 *Let $n \geq 1$ and consider a sequence $a_1, a_2, \ldots, a_{n+1}$ of $n+1$ elements from the set $\{1, 2, \ldots, 2n\}$. Then there are two distinct indices i and j such that a_i divides a_j or a_j divides a_i.*

Proof. For each i with $1 \leq i \leq n + 1$, write

$$a_i = 2^{k_i} \cdot q_i,$$

where $k_i \geq 0$ is an integer and q_i is an odd integer. For example,

- if $a_i = 48$, then $k_i = 4$ and $q_i = 3$, because $48 = 2^4 \cdot 3$,
- if $a_i = 1$, then $k_i = 0$ and $q_i = 1$, because $1 = 2^0 \cdot 1$,
- if $a_i = 7$, then $k_i = 0$ and $q_i = 7$, because $7 = 2^0 \cdot 7$.

Consider the sequence $q_1, q_2, \ldots, q_{n+1}$ of $n+1$ integers. Each of these numbers is an odd integer that belongs to the set

$$\{1, 3, 5, \ldots, 2n - 1\} = \{2m - 1 : m = 1, 2, \ldots, n\}.$$

Since this set has size n, the Pigeonhole Principle implies that there must be two numbers in the sequence $q_1, q_2, \ldots, q_{n+1}$ that are equal. In other words, there are two distinct indices i and j such that $q_i = q_j$. It follows that

$$\frac{a_i}{a_j} = \frac{2^{k_i} \cdot q_i}{2^{k_j} \cdot q_j} = 2^{k_i - k_j}.$$

Thus, if $k_i \geq k_j$, then a_j divides a_i. Otherwise, $k_i < k_j$, and a_i divides a_j. ■

3.10.3 Long Sequences Contain Long Monotone Subsequences

Let $n = 3$, and consider the sequence $20, 10, 9, 7, 11, 2, 21, 1, 20, 31$ of $10 = n^2 + 1$ numbers. This sequence contains an increasing subsequence of length $4 = n + 1$, namely $10, 11, 21, 31$. The following theorem states this result for arbitrary values of n.

Theorem 3.10.2 *Let $n \geq 1$ be an integer. Every sequence of $n^2 + 1$ distinct real numbers contains a subsequence of length $n + 1$ that is either increasing or decreasing.*

Proof. Let $a_1, a_2, \ldots, a_{n^2+1}$ be an arbitrary sequence of $n^2 + 1$ distinct real numbers. For each i with $1 \leq i \leq n^2 + 1$, let inc_i denote the length of the longest increasing subsequence that starts at a_i, and let dec_i denote the length of the longest decreasing subsequence that starts at a_i.

Using this notation, the claim in the theorem can be formulated as follows: There is an index i such that $inc_i \geq n + 1$ or $dec_i \geq n + 1$.

We will prove the claim by contradiction. Thus, we assume that $inc_i \leq n$ and $dec_i \leq n$ for all i with $1 \leq i \leq n^2 + 1$.

Consider the set

$$B = \{(b, c) : 1 \leq b \leq n, 1 \leq c \leq n\},$$

and think of the elements of B as being boxes. For each i with $1 \leq i \leq n^2+1$, the pair (inc_i, dec_i) is an element of B. Thus, we have $n^2 + 1$ elements

(inc_i, dec_i), which are placed in the n^2 boxes of B. By the Pigeonhole Principle, there must be a box that contains two (or more) elements. In other words, there exist two integers i and j such that $i < j$ and

$$(inc_i, dec_i) = (inc_j, dec_j).$$

Recall that the elements in the sequence are distinct. Hence, $a_i \neq a_j$. We consider two cases.

First assume that $a_i < a_j$. Then the length of the longest increasing subsequence starting at a_i must be at least $1+inc_j$, because we can append a_i to the longest increasing subsequence starting at a_j. Therefore, $inc_i \neq inc_j$, which is a contradiction.

The second case is when $a_i > a_j$. Then the length of the longest decreasing subsequence starting at a_i must be at least $1+dec_j$, because we can append a_i to the longest decreasing subsequence starting at a_j. Therefore, $dec_i \neq dec_j$, which is again a contradiction. ∎

3.10.4 There are Infinitely Many Primes

As a final application of the Pigeonhole Principle, we prove the following result:

Theorem 3.10.3 *There are infinitely many prime numbers.*

Proof. The proof is by contradiction. Thus, we assume that there are, say, k prime numbers, and denote them by

$$2 = p_1 < p_2 < \cdots < p_k.$$

Note that k is a fixed integer. Since

$$\lim_{n\to\infty} \frac{2^n}{(n+1)^k} = \infty,$$

we can choose an integer n such that

$$2^n > (n+1)^k.$$

Define the function

$$f : \{1, 2, \ldots, 2^n\} \to \mathbb{N}^k$$

as follows: For any integer x with $1 \leq x \leq 2^n$, consider its prime factorization

$$x = p_1^{m_1} \cdot p_2^{m_2} \cdots p_k^{m_k}.$$

We define

$$f(x) = (m_1, m_2, \ldots, m_k).$$

Since

$$\begin{aligned} m_i &\leq m_1 + m_2 + \cdots + m_k \\ &\leq m_1 \log p_1 + m_2 \log p_2 + \cdots + m_k \log p_k \\ &= \log (p_1^{m_1} \cdot p_2^{m_2} \cdots p_k^{m_k}) \\ &= \log x \\ &\leq n, \end{aligned}$$

it follows that

$$f(x) \in \{0, 1, 2, \ldots, n\}^k.$$

Thus, f is a function

$$f : \{1, 2, \ldots, 2^n\} \to \{0, 1, 2, \ldots, n\}^k.$$

It is easy to see that this function is one-to-one. The set on the left-hand side has size 2^n, whereas the set on the right-hand side has size $(n+1)^k$. It then follows from the Pigeonhole Principle that

$$(n+1)^k \geq 2^n,$$

which contradicts our choice for n. ∎

3.11 Exercises

3.1 A licence plate number consists of a sequence of four uppercase letters followed by three digits. How many licence plate numbers are there?

3.2 A multiple-choice exam consists of 100 questions. Each question has four possible answers a, b, c, and d. How many ways are there to answer the 100 questions (assuming that each question is answered)?

3.3 For each of the following 7 cases, determine how many strings of 8 uppercase letters there are.

- Letters can be repeated.
- No letter can be repeated.
- The strings start with PQ (in this order) and letters can be repeated.
- The strings start with PQ (in this order) and no letter can be repeated.
- The strings start and end with PQ (in this order) and letters can be repeated.
- The strings start with XYZ (in this order), end with QP (in this order), and letters can be repeated.
- The strings start with XYZ (in this order) or end with QP (in this order), and letters can be repeated.

3.4 The Carleton Computer Science Society has a Board of Directors consisting of one president, one vice-president, one secretary, one treasurer, and a three-person party committee (whose main responsibility is to buy beer for the other four board members). The entire board consists of seven distinct students. If there are $n \geq 7$ students in Carleton's Computer Science program, how many ways are there to choose a Board of Directors?

3.5 There are $n \geq 4$ students in Carleton's Computer Science program. The Carleton Computer Science Society has a Board of Directors consisting of one president and three vice-presidents. The entire board consists of four distinct students. Prove that

$$n\binom{n-1}{3} = (n-3)\binom{n}{3},$$

by counting, in two different ways, the number of ways to choose a Board of Directors.

3.6 In how many ways can you paint 200 chairs, if 33 of them must be painted red, 66 of them must be painted blue, and 101 of them must be painted green?

3.7 Let A be a set of size m, let B be a set of size n, and assume that $n \geq m \geq 1$. How many functions $f : A \to B$ are there that are *not* one-to-one?

3.8 How many bitstrings of length 8 are there that contain at least 4 consecutive 0s or at least 4 consecutive 1s?

3.9 How many bitstrings of length 77 are there that start with 010 (i.e., have 010 at positions 1, 2, and 3) or have 101 at positions 2, 3, and 4, or have 010 at positions 3, 4, and 5?

3.10 Let m and n be integers with $m \geq n \geq 1$. How many ways are there to place m books on n shelves, if there must be at least one book on each shelf? As in Section 3.1.3, the order on each shelf matters.

3.11 You are given m distinct books $B_1, B_2, \ldots, B_m$ and n *identical* blocks of wood. How many ways are there to arrange these books and blocks in a straight line?

For example, if $m = 5$ and $n = 3$, then three possible arrangements are (W stands for a block of wood)

$$WB_3B_1WB_5B_4WB_2,$$

$$WB_1B_3WB_5B_4WB_2,$$

and

$$B_5WB_3B_1WWB_2B_4.$$

3.12 Let $n \geq 1$ be an integer and consider n boys and n girls. For each of the following three cases, determine how many ways there are to arrange these $2n$ people on a straight line (the order on the line matters):

- All boys stand next to each other and all girls stand next to each other.
- All girls stand next to each other.
- Boys and girls alternate.

3.13 Consider strings consisting of 12 characters, each character being a, b, or c. Such a string is called *valid* if at least one of the characters is missing. For example, *abababababab* is a valid string, whereas *abababacabab* is not a valid string. How many valid strings are there?

3.14 A password consists of 100 characters, each character being a digit or a lowercase letter. A password must contain at least two digits. How many passwords are there?

3.15 A password consists of 100 characters, each character being a digit, a lowercase letter, or an uppercase letter. A password must contain at least one digit, at least one lowercase letter, and at least one uppercase letter. How many passwords are there?
Hint: Recall De Morgan's Law

$$A \cap B \cap C = \overline{\overline{A} \cup \overline{B} \cup \overline{C}}.$$

3.16 In a group of 20 people,

- 6 are blond,
- 7 have green eyes,
- 11 are not blond and do not have green eyes.

How many people are blond and have green eyes?

3.17 Use Pascal's Identity (Theorem 3.7.2) to prove Newton's Binomial Theorem (Theorem 3.6.5) by induction.

3.18 Determine the coefficient of $x^{111}y^{444}$ in the expansion of

$$(-17x + 71y)^{555}.$$

3.19 Let $n \geq 1$ be an integer. Use Newton's Binomial Theorem (Theorem 3.6.5) to prove that

$$\sum_{k=1}^{n} \binom{n}{k} 10^k \cdot 26^{n-k} = 36^n - 26^n. \tag{3.5}$$

In the rest of this exercise, you will give a combinatorial proof of this identity.

Consider passwords consisting of n characters, each character being a digit or a lowercase letter. A password must contain at least one digit.

- Use the Complement Rule of Section 3.3 to show that the number of passwords is equal to $36^n - 26^n$.

- Let k be an integer with $1 \leq k \leq n$. Prove that the number of passwords with exactly k digits is equal to $\binom{n}{k}10^k \cdot 26^{n-k}$.

- Explain why the above two parts imply the identity in (3.5).

3.20 Use Newton's Binomial Theorem (Theorem 3.6.5) to prove that for every integer $n \geq 1$,

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n.$$

In the rest of this exercise, you will give a combinatorial proof of this identity.

Let $A = \{1, 2, 3, \ldots, n\}$ and $B = \{a, b, c\}$. According to Theorem 3.1.2, the number of functions $f : A \to B$ is equal to 3^n.

- Consider a fixed integer k with $0 \leq k \leq n$ and a fixed subset S of A having size k. How many functions $f : A \to B$ are there for which
  - for all $x \in S$, $f(x) \in \{a, b\}$, and
  - for all $x \in A \setminus S$, $f(x) = c$.

- Argue that the summation $\sum_{k=0}^{n} \binom{n}{k} 2^k$ counts all functions $f : A \to B$.

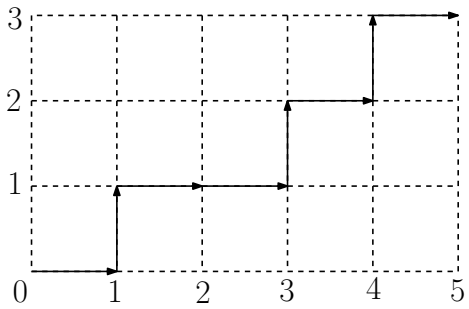
3.21 Let $n \geq 66$ be an integer and consider the set $S = \{1, 2, \ldots, n\}$.

- Let k be an integer with $66 \leq k \leq n$. How many 66-element subsets of S are there whose largest element is equal to k?

- Use the result in the first part to prove that

$$\sum_{k=66}^{n} \binom{k-1}{65} = \binom{n}{66}.$$

3.22 Let $m \geq 1$ and $n \geq 1$ be integers. Consider a rectangle whose horizontal side has length m and whose vertical side has length n. A path from the bottom-left corner to the top-right corner is called *valid*, if in each step, it either goes one unit to the right or one unit upwards. In the example below, you see a valid path for the case when $m = 5$ and $n = 3$.



How many valid paths are there?

3.23 How many different strings can be obtained by reordering the letters of the word MississippiMills. (This is a town close to Ottawa. James Naismith, the inventor of basketball, was born there.)

3.24 In Theorems 3.9.1 and 3.9.2, we have seen how many solutions (in nonnegative integers) there are for equations of the type

$$x_1 + x_2 + \cdots + x_k = n$$

and inequalities of the type

$$x_1 + x_2 + \cdots + x_k \leq n.$$

Use this to prove the following identity:

$$\sum_{i=0}^{n} \binom{i+k-1}{k-1} = \binom{n+k}{k}.$$

3.25 Let n and k be integers with $n \geq k \geq 1$. How many solutions are there to the equation

$$x_1 + x_2 + \cdots + x_k = n,$$

where $x_1 \geq 1$, $x_2 \geq 1$, ..., $x_k \geq 1$ are integers?
Hint: In Theorem 3.9.1, we have seen the answer if $x_1 \geq 0$, $x_2 \geq 0$, ..., $x_k \geq 0$.

3.26 Let $n \geq 1$ be an integer. Use the Pigeonhole Principle to prove that in any set of $n+1$ integers from $\{1, 2, \ldots, 2n\}$, there are two integers that are consecutive (i.e., differ by one).

3.27 Consider five points in a square with sides of length one. Use the Pigeonhole Principle to prove that there are two of these points having distance at most $1/\sqrt{2}$.

3.28 Let S be a set of 90 positive integers, each one having at most 25 digits in decimal notation. Use the Pigeonhole Principle to prove that there are two different subsets A and B of S that have the same sum, i.e.,

$$\sum_{x \in A} x = \sum_{x \in B} x.$$

Chapter 4

Recursion

> In order to understand recursion, you must first understand recursion.

Recursion is the concept where an object (such as a function, a set, or an algorithm) is defined in the following way:

- There are one or more base cases.
- There are one or more rules that define an object in terms of "smaller" objects that have already been defined.

In this chapter, we will see several examples of such recursive definitions and how to use them to solve counting problems.

4.1 Recursive Functions

Recall that $\mathbb{N} = \{0, 1, 2, \ldots\}$ denotes the set of natural numbers. Consider the following recursive definition of a function $f : \mathbb{N} \to \mathbb{N}$:

$$\begin{aligned} f(0) &= 3, \\ f(n) &= 2 \cdot f(n-1) + 3, \text{ if } n \geq 1. \end{aligned}$$

These two rules indeed *define* a function, because $f(0)$ is uniquely defined and for any integer $n \geq 1$, if $f(n-1)$ is uniquely defined, then $f(n)$ is also uniquely defined, because it is equal to $2 \cdot f(n-1)+3$. Therefore, by induction, for any natural number n, the function value $f(n)$ is uniquely defined. We can obtain the values $f(n)$ in the following way:

- We are given that $f(0) = 3$.
- If we apply the recursive rule with $n = 1$, then we get

$$f(1) = 2 \cdot f(0) + 3 = 2 \cdot 3 + 3 = 9.$$

- If we apply the recursive rule with $n = 2$, then we get

$$f(2) = 2 \cdot f(1) + 3 = 2 \cdot 9 + 3 = 21.$$

- If we apply the recursive rule with $n = 3$, then we get

$$f(3) = 2 \cdot f(2) + 3 = 2 \cdot 21 + 3 = 45.$$

- If we apply the recursive rule with $n = 4$, then we get

$$f(4) = 2 \cdot f(3) + 3 = 2 \cdot 45 + 3 = 93.$$

Can we "solve" this recurrence? That is, can we express $f(n)$ in terms of n only? By looking at these values, you may see a pattern, i.e., you may guess that for each $n \geq 0$,

$$f(n) = 3 \cdot 2^{n+1} - 3. \tag{4.1}$$

We prove by induction that this is correct: If $n = 0$, then $f(n) = f(0) = 3$ and $3 \cdot 2^{n+1} - 3 = 3 \cdot 2^{0+1} - 3 = 3$. Thus, (4.1) is true for $n = 0$. Let $n \geq 1$ and assume that (4.1) is true for $n - 1$, i.e., assume that

$$f(n-1) = 3 \cdot 2^n - 3.$$

Then

$$\begin{aligned} f(n) &= 2 \cdot f(n-1) + 3 \\ &= 2\left(3 \cdot 2^n - 3\right) + 3 \\ &= 3 \cdot 2^{n+1} - 3. \end{aligned}$$

Thus, we have proved by induction that (4.1) holds for all integers $n \geq 0$.

A recursive definition of factorials: Consider the following recursive definition of a function $g : \mathbb{N} \to \mathbb{N}$:

$$\begin{aligned} g(0) &= 1, \\ g(n) &= n \cdot g(n-1), \text{ if } n \geq 1. \end{aligned}$$

As in the previous example, a simple induction proof shows that these rules uniquely define the value $g(n)$ for each $n \geq 0$. We leave it to the reader to verify that g is the factorial function, i.e., $g(n) = n!$ for each $n \geq 0$.

A recursive definition of binomial coefficients: Consider the following recursive definition of a function $B : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ with two variables:

$$\begin{aligned} B(n,0) &= 1, \text{ if } n \geq 0, \\ B(n,n) &= 1, \text{ if } n \geq 0, \\ B(n,k) &= B(n-1,k-1) + B(n-1,k), \text{ if } n \geq 2 \text{ and } 1 \leq k \leq n-1. \end{aligned}$$

The recursive rule has the same form as Pascal's Identity in Theorem 3.7.2. Also, the first base case shows that $B(n,0) = 1 = \binom{n}{0}$, whereas the second base case shows that $B(n,n) = 1 = \binom{n}{n}$. From this, it can be shown by induction that $B(n,k) = \binom{n}{k}$ for all n and k with $0 \leq k \leq n$.

4.2 Fibonacci Numbers

I'll have an order of the Fibonachos.

The *Fibonacci numbers* are defined using the following rules:

$$\begin{aligned} f_0 &= 0, \\ f_1 &= 1, \\ f_n &= f_{n-1} + f_{n-2}, \text{ if } n \geq 2. \end{aligned}$$

In words, there are two base cases (i.e., 0 and 1) and each next element in the sequence is the sum of the previous two elements. This gives the sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

The following theorem states that we can "solve" this recurrence. That is, we can express the n-th Fibonacci number f_n in a non-recursive way, i.e., without using any other Fibonacci numbers.

Theorem 4.2.1 *Let $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$ be the two solutions of the quadratic equation $x^2 = x + 1$. Then, for all $n \geq 0$, we have*

$$f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}.$$

Proof. We prove the claim by induction on n. There are two base cases[1]:

- Both f_0 and $\frac{\varphi^0 - \psi^0}{\sqrt{5}}$ are equal to 0.
- Both f_1 and $\frac{\varphi^1 - \psi^1}{\sqrt{5}}$ are equal to 1.

Let $n \geq 2$ and assume that the claim is true for $n-2$ and $n-1$. In other words, assume that

$$f_{n-2} = \frac{\varphi^{n-2} - \psi^{n-2}}{\sqrt{5}}$$

and

$$f_{n-1} = \frac{\varphi^{n-1} - \psi^{n-1}}{\sqrt{5}}.$$

We have to prove that the claim is true for n as well. Using the definition of f_n, the two assumptions, and the identities $\varphi^2 = \varphi + 1$ and $\psi^2 = \psi + 1$, we get

$$\begin{aligned} f_n &= f_{n-1} + f_{n-2} \\ &= \frac{\varphi^{n-1} - \psi^{n-1}}{\sqrt{5}} + \frac{\varphi^{n-2} - \psi^{n-2}}{\sqrt{5}} \\ &= \frac{\varphi^{n-2}(\varphi + 1)}{\sqrt{5}} - \frac{\psi^{n-2}(\psi + 1)}{\sqrt{5}} \\ &= \frac{\varphi^{n-2} \cdot \varphi^2}{\sqrt{5}} - \frac{\psi^{n-2} \cdot \psi^2}{\sqrt{5}} \\ &= \frac{\varphi^n - \psi^n}{\sqrt{5}}. \end{aligned}$$

∎

[1] Do you see why there are two base cases?

4.2.1 Counting 00-Free Bitstrings

Consider finite bitstrings that do not contain 00. Examples of such bitstrings are 10, 010, 0101010101, and 11110111. For any integer $n \geq 2$, how many such strings are there having length n? Since we do not know the answer yet, we introduce a variable B_n, one for each $n \geq 2$, for the number of such strings. Thus,

- B_n denotes the number of bitstrings of length n that do not contain 00.

There are four bitstrings of length 2:

$$00, 10, 01, 11.$$

Since three of them do not contain 00, it follows that $B_2 = 3$. Similarly, there are eight bitstrings of length 3:

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

Since five of them do not contain 00, we have $B_3 = 5$.

Let $n \geq 4$. We are going to express B_n in terms of the previous two values B_{n-1} and B_{n-2}. This, together with the two base cases $B_2 = 3$ and $B_3 = 5$, will give a recurrence for the entire sequence.

Consider a matrix that contains all bitstrings of length n that do not contain 00, one string per row. Since there are B_n many such strings, the matrix has B_n many rows. Also, the matrix has n columns, because the strings have length n.

We rearrange the rows of the matrix such that all strings in the *top part* start with 1 and all strings in the *bottom part* start with 0.

- How many rows are there in the top part? Any string in the top part starts with 1 and is followed by a bitstring of length $n-1$ that does not contain 00. Thus, if we take the rows in the top part and delete the first bit from each row, then we obtain all bitstrings of length $n-1$ that do not contain 00. Since there are B_{n-1} many such strings of length $n-1$, it follows that the top part of the matrix consists of B_{n-1} many rows.

- How many rows are there in the bottom part? Any string in the bottom part starts with 0. Since the string does not contain 00, the second bit

must be 1. After these first two bits, we have a bitstring of length $n-2$ that does not contain 00. Thus, if we take the rows in the bottom part and delete the first two bits from each row, then we obtain all bitstrings of length $n-2$ that do not contain 00. Since there are B_{n-2} many such strings of length $n-2$, it follows that the bottom part of the matrix consists of B_{n-2} many rows.

Thus, on the one hand, the matrix has B_n many rows. On the other hand, this matrix has $B_{n-1} + B_{n-2}$ many rows. Therefore, we have $B_n = B_{n-1} + B_{n-2}$.

To summarize, we have proved that the values B_n, for $n \geq 2$, satisfy the following recurrence:

$$\begin{aligned} B_2 &= 3, \\ B_3 &= 5, \\ B_n &= B_{n-1} + B_{n-2}, \text{ if } n \geq 4. \end{aligned}$$

This recurrence is the same as the one for the Fibonacci numbers, except that the two base cases are different. The sequence B_n, $n \geq 2$, consists of the integers

$$3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

We obtain this sequence by removing the first four elements (i.e., f_0, f_1, f_2, and f_3) from the Fibonacci sequence. We leave it to the reader to verify (using induction) that for all $n \geq 2$,

$$B_n = f_{n+2}.$$

4.3 A Recursively Defined Set

Consider the set S that is defined by the following two rules:

- 5 is an element of the set S.
- If x and y are elements of the set S, then $x - y$ is also an element of the set S.

Thus, if we already know that x and y belong to the set S, then the second rule gives us a new element, i.e., $x - y$, that also belongs to S.

Can we give a simple description of the set S? We are going to use the rules to obtain some elements of S. From these examples, we then hope to see a pattern from which we guess the simple description of S. The final step consists of proving that our guess is correct.

- We are given that 5 is an element of S.
- Applying the rule with $x = 5$ and $y = 5$ implies that $x - y = 0$ is also an element of S.
- Applying the rule with $x = 0$ and $y = 5$ implies that $x - y = -5$ is also an element of S.
- Applying the rule with $x = 5$ and $y = -5$ implies that $x - y = 10$ is also an element of S.
- Applying the rule with $x = 0$ and $y = 10$ implies that $x - y = -10$ is also an element of S.
- Applying the rule with $x = 5$ and $y = -10$ implies that $x - y = 15$ is also an element of S.
- Applying the rule with $x = 0$ and $y = 15$ implies that $x - y = -15$ is also an element of S.

Thus, we have obtained the following elements of S:

$$-15, -10, -5, 0, 5, 10, 15$$

Since there is clearly a pattern, it is natural to guess that

$$S = \{5n : n \in \mathbb{Z}\}, \tag{4.2}$$

where $\mathbb{Z}$ is the set of all (positive and negative) integers, including 0. To prove that this is correct, we will first prove that the set on the left-hand side is a subset of the set on the right-hand side. Then we prove that the set on the right-hand side is a subset of the set on the left-hand side.

We start by proving that

$$S \subseteq \{5n : n \in \mathbb{Z}\},$$

which is equivalent to proving that

$$\text{every element of } S \text{ is a multiple of 5.} \tag{4.3}$$

How do we prove this? The set S is defined using a base case and a recursive rule. The only way to obtain an element of S is by starting with the base case and then applying the recursive rule a finite number of times. Therefore, the following will prove that (4.3) holds:

- 5 is a multiple of 5.
- Let x and y be two elements of S and assume that they are both multiples of 5. Then $x - y$ (which is the "next" element of S) is also a multiple of 5.

Next we prove that

$$\{5n : n \in \mathbb{Z}\} \subseteq S.$$

We will do this by proving that for all $n \geq 0$,

$$5n \in S \text{ and } -5n \in S. \tag{4.4}$$

The proof is by induction on n. For the base case, i.e., when $n = 0$, we observe that, from the definition of S, $x = 5$ and $y = 5$ are in S and, therefore, $x - y = 0$ is also in S. Therefore, (4.4) is true for $n = 0$.

Let $n \geq 0$ and assume that (4.4) is true for n, i.e., assume that

$$5n \in S \text{ and } -5n \in S.$$

We have to show that (4.4) is also true for $n + 1$, i.e.,

$$5(n+1) \in S \text{ and } -5(n+1) \in S.$$

- It follows from the definition of S and our assumption that both $x = 5$ and $y = -5n$ are in S. Therefore, $x - y = 5(n + 1)$ is also in S.
- It follows from the definition of S and our assumption that both $x = -5n$ and $y = 5$ are in S. Therefore, $x - y = -5(n + 1)$ is also in S.

Thus, we have shown by induction that (4.4) holds for all $n \geq 0$.

Since we have shown that both (4.3) and (4.4) hold, we conclude that (4.2) holds as well. In other words, we have indeed obtained a simple description of the set S: It is the set of all multiples of 5.

4.4 A Gossip Problem

Let $n \geq 4$ be an integer and consider a group $P_1, P_2, \ldots, P_n$ of n people. Assume that each person P_i knows some scandal S_i that nobody else knows. For any i and j, if person P_i makes a phone call with person P_j, they exchange

the scandals they know at that moment, i.e., P_i tells all scandals he knows to P_j, and P_j tells all scandals he knows to P_i. How many phone calls are needed until each of the n people knows all n scandals?

An obvious solution is that each pair of people in the group makes one phone call. At the end, each person knows all scandals. The number of phone calls is

$$\binom{n}{2} = \frac{n(n-1)}{2},$$

which is quadratic in the number n of people. We will see below that only a linear number of phone calls are needed.

Let us first consider the case when $n = 4$. At the start, each person P_i only knows the scandal S_i, which we visualize in the following table:

P_1	P_2	P_3	P_4
S_1	S_2	S_3	S_4

Consider the following sequence of phone calls:

1. P_1 calls P_2. After this phone call, the table looks as follows:

P_1	P_2	P_3	P_4
S_1S_2	S_1S_2	S_3	S_4

2. P_3 calls P_4. After this phone call, the table looks as follows:

P_1	P_2	P_3	P_4
S_1S_2	S_1S_2	S_3S_4	S_3S_4

3. P_1 calls P_3. After this phone call, the table looks as follows:

P_1	P_2	P_3	P_4
$S_1S_2S_3S_4$	S_1S_2	$S_1S_2S_3S_4$	S_3S_4

4. P_2 calls P_4. After this phone call, the table looks as follows:

P_1	P_2	P_3	P_4
$S_1S_2S_3S_4$	$S_1S_2S_3S_4$	$S_1S_2S_3S_4$	$S_1S_2S_3S_4$

We see that after four phone calls, each person knows all four scandals. Observe that the number of phone calls is $\binom{4}{2} = 6$ if we would have used the obvious solution mentioned above.

Thus, we now have an algorithm that schedules the phone calls for groups of four people. Below, we will extend this "base case" to a recursive algorithm that schedules the phone calls for any group of $n \geq 4$ people. The approach is as follows:

- We assume that we know how to schedule the phone calls for groups of $n - 1$ people.
- We use this assumption to schedule the phone calls for groups of n people.

Let us see how this is done.

- At the start, P_1 knows S_1, P_2 knows S_2, ..., P_n knows S_n.
- P_{n-1} calls P_n. After this phone call, P_1 knows S_1, P_2 knows S_2, ..., P_{n-2} knows S_{n-2}, and both P_{n-1} and P_n know S_{n-1} and S_n.
- Consider S_{n-1} and S_n to be one scandal S'_{n-1}.
- Schedule the phone calls for the group $P_1, P_2, \ldots, P_{n-1}$ of $n-1$ people, using the scandals $S_1, S_2, \ldots, S_{n-2}, S'_{n-1}$. (We have assumed that we know how to do this!) At the end, each of $P_1, P_2, \ldots, P_{n-1}$ knows all scandals $S_1, S_2, \ldots, S_n$.
- At this moment, P_n only knows S_{n-1} and S_n. Therefore, P_{n-1} again calls P_n and tells him all scandals $S_1, S_2, \ldots, S_{n-2}$.

Below, you see this recursive algorithm in pseudocode.

Algorithm GOSSIP(n)**:**

// $n \geq 4$, this algorithm schedules phone calls for $P_1, P_2, \ldots, P_n$
if $n = 4$
then P_1 calls P_2;
P_3 calls P_4;
P_1 calls P_3;
P_2 calls P_4
else P_{n-1} calls P_n;
GOSSIP($n-1$);
P_{n-1} calls P_n
endif

What is the number of phone calls made when running algorithm GOSSIP(n)? Since we do not know the answer yet, we introduce a variable $C(n)$ to denote this number. It follows from the pseudocode that

$$C(4) = 4.$$

Let $n \geq 5$. Algorithm GOSSIP(n) starts and ends with the same phone call: P_{n-1} calls P_n. In between, it runs algorithm GOSSIP($n-1$), during which, by definition, $C(n-1)$ phone calls are made. It follows that

$$C(n) = 2 + C(n-1) \text{ for } n \geq 5.$$

Thus, we have obtained a recurrence for the numbers $C(n)$. The first few numbers in the sequence are

$$\begin{array}{lllllll} C(4) & = & 4, & & & & \\ C(5) & = & 2 + C(4) & = & 2 + 4 & = & 6, \\ C(6) & = & 2 + C(5) & = & 2 + 6 & = & 8, \\ C(7) & = & 2 + C(6) & = & 2 + 8 & = & 10. \end{array}$$

From this, we guess that

$$C(n) = 2n - 4 \text{ for } n \geq 4.$$

We can easily prove by induction that our guess is correct. Indeed, since both $C(4)$ and $2 \cdot 4 - 4$ are equal to 4, the claim is true for $n = 4$. If $n \geq 5$ and $C(n-1) = 2(n-1) - 4$, then

$$C(n) = 2 + C(n-1) = 2 + (2(n-1) - 4) = 2n - 4.$$

This shows that $C(n) = 2n - 4$ for all $n \geq 4$.

It can be shown that algorithm GOSSIP is optimal: Any algorithm that schedules phone calls for $n \geq 4$ people must make at least $2n-4$ phone calls.

You may wonder why the base case for algorithm GOSSIP(n) is when $n = 4$. You will find the reason in Exercise 4.17.

4.5 The Merge-Sort Algorithm

MERGESORT is a recursive sorting algorithm that works as follows. To sort the sequence $a_1, a_2, \ldots, a_n$ of numbers,

- it recursively sorts the sequence $a_1, a_2, \ldots, a_m$, where $m = \lfloor n/2 \rfloor$, and stores the sorted sequence in a list L_1,
- it recursively sorts the sequence $a_{m+1}, a_{m+2}, \ldots, a_n$ and stores the sorted sequence in a list L_2,
- it merges the two sorted lists L_1 and L_2 into one sorted list.

Below, you see this recursive algorithm in pseudocode.

Algorithm MERGESORT(L, n)**:**

```
// L is a list of n ≥ 0 numbers
if n ≥ 2
then m = ⌊n/2⌋;
     L1 = list containing the first m elements of L;
     L2 = list containing the last n − m elements of L;
     L1 = MERGESORT(L1, m);
     L2 = MERGESORT(L2, n − m);
     L = MERGE(L1, L2)
endif;
return L
```

We still have to specify algorithm MERGE(L_1, L_2). Of course, this algorithm uses the fact that both L_1 and L_2 are sorted lists. The task is to merge them into one sorted list. This is done in the following way. Initialize an empty list L. (At the end, this list will contain the final sorted sequence.)

- Let x be the first element of L_1 and let y be the first element of L_2.

- If $x \leq y$, then remove x from L_1 and append it to L (i.e., add x at the end of L).
- Otherwise (i.e., if $x > y$), remove y from L_2 and append it to L.

Repeat these steps until one of L_1 and L_2 is empty. If L_1 is empty, then append L_2 to L. Otherwise, append L_1 to L. Here is the algorithm in pseudocode:

Algorithm MERGE(L_1, L_2)**:**

```
// L1 and L2 are sorted lists
L = empty list;
while L1 is not empty and L2 is not empty
do x = first element of L1;
    y = first element of L2;
    if x ≤ y
    then remove x from L1;
           append x to L
    else remove y from L2;
          append y to L
    endif
endwhile;
if L1 is non-empty
then append L1 to L
else append L2 to L
endif;
return L
```

4.5.1 Correctness of Algorithm MERGESORT

I hope you are convinced that the output L of algorithm MERGE(L_1, L_2) is a sorted list that contains all elements of L_1 and L_2 (and no other elements). How do we prove that algorithm MERGESORT(L, n) is correct, i.e., correctly sorts the elements in any list L of n numbers? Since the algorithm is recursive, we prove this by induction.

The two base cases are when $n = 0$ or $n = 1$. It follows from the pseudocode for MERGESORT(L, n) that it simply returns the input list L,

which is obviously sorted.

Let $n \geq 2$ and assume that for any integer k with $0 \leq k < n$ and for any list L' of k numbers, algorithm MergeSort(L', k) returns a list containing the elements of L' in sorted order. Let L be a list of n numbers. By going through the pseudocode for MergeSort(L, n), we observe the following:

- The recursive call MergeSort(L_1, m) is on a list with less than n numbers. Therefore, by the induction hypothesis, its output, which is the list L_1, is sorted.
- The recursive call MergeSort$(L_2, n-m)$ is on a list with less than n numbers. Again by the induction hypothesis, its output, which is the list L_2, is sorted.
- Algorithm Merge$(L_1, L_2,)$ gets as input the two sorted lists L_1 and L_2, and returns a list L. Since algorithm Merge is correct, it then follows that L is a sorted list.

It follows that the final list L, which is returned by algorithm MergeSort, is sorted.

This proves the correctness of algorithm MergeSort(L, n) for any integer $n \geq 0$ and any list L of n numbers.

4.5.2 Running Time of Algorithm MergeSort

We now analyze the running time of algorithm MergeSort. It follows from the pseudocode that, when running this algorithm together with its recursive calls, several calls are made to algorithm Merge. We are going to count the total number of *comparisons* that are made. That is, we will determine the total number of times that the line "**if** $x \leq y$" in algorithm Merge is executed when running algorithm MergeSort(L, n).

We first observe that the number of comparisons made by algorithm Merge(L_1, L_2) is at most $|L_1| + |L_2|$.

Let n be an integer and assume for simplicity that n is a power of two, i.e., $n = 2^k$ for some integer $k \geq 0$. We define $T(n)$ to be the maximum number of comparisons made when running algorithm MergeSort(L, n) on any input list L of n numbers. Note that we include in $T(n)$ all comparisons that are made during all calls to Merge that are part of all recursive calls that are generated when running MergeSort(L, n).

Consider a list L of n numbers. It follows from the pseudocode for MergeSort(L, n) that

$$T(1) = 0.$$

Let $n \geq 2$ and consider again the pseudocode for MergeSort(L, n). Which parts of the algorithm make comparisons between input elements?

- The call MergeSort(L_1, m) is a recursive call on a list of $m = n/2$ numbers. By definition, the total number of comparisons made in this call (together with all its recursive subcalls) is at most $T(n/2)$.

- The call MergeSort$(L_2, n-m)$ is a recursive call on a list of $n-m = n/2$ numbers. By definition, the total number of comparisons made in this call (together with all its recursive subcalls) is at most $T(n/2)$.

- Finally, algorithm MergeSort(L, n) calls the non-recursive algorithm Merge(L_1, L_2). We have seen above that the number of comparisons made in this call is at most $|L_1| + |L_2| = n$.

By adding the number of comparisons, we get

$$T(n) \leq T(n/2) + T(n/2) + n = 2 \cdot T(n/2) + n.$$

Thus, we obtain the following recurrence:

$$\begin{aligned} T(1) &= 0, \\ T(n) &\leq 2 \cdot T(n/2) + n, \text{ if } n \geq 2 \text{ is a power of 2.} \end{aligned} \tag{4.5}$$

Our goal was to determine $T(n)$, but at this moment, we only have a recurrence for this function. We will solve the recurrence using a technique called *unfolding*:

Recall that we assume that $n = 2^k$ for some integer $k \geq 0$. Let n be a large integer. We know from (4.5) that

$$T(n) \leq 2 \cdot T(n/2) + n.$$

If we replace n by $n/2$ in (4.5), which is a valid thing to do, we get

$$T(n/2) \leq 2 \cdot T(n/2^2) + n/2.$$

By combining these two inequalities, we get

$$\begin{aligned} T(n) &\leq 2 \cdot T(n/2) + n \\ &\leq 2\left(2 \cdot T(n/2^2) + n/2\right) + n \\ &= 2^2 \cdot T(n/2^2) + 2n. \end{aligned}$$

Let us repeat this: Replacing n by $n/2^2$ in (4.5) gives

$$T(n/2^2) \leq 2 \cdot T(n/2^3) + n/2^2.$$

By substituting this into the inequality for $T(n)$, we get

$$\begin{aligned} T(n) &\leq 2^2 \cdot T(n/2^2) + 2n \\ &\leq 2^2\left(2 \cdot T(n/2^3) + n/2^2\right) + 2n \\ &= 2^3 \cdot T(n/2^3) + 3n. \end{aligned}$$

In the next step, we replace n by $n/2^3$ in (4.5), which gives

$$T(n/2^3) \leq 2 \cdot T(n/2^4) + n/2^3.$$

By substituting this into the inequality for $T(n)$, we get

$$\begin{aligned} T(n) &\leq 2^3 \cdot T(n/2^3) + 3n \\ &\leq 2^3\left(2 \cdot T(n/2^4) + n/2^3\right) + 3n \\ &= 2^4 \cdot T(n/2^4) + 4n. \end{aligned}$$

At this moment, you will see the pattern and, at the end, we get the inequality

$$T(n) \leq 2^k \cdot T(n/2^k) + kn.$$

Since $n = 2^k$, we have $T(n/2^k) = T(1)$, which is 0 from the base case of the recurrence. Also, $n = 2^k$ implies that $k = \log n$. We conclude that

$$T(n) \leq n \cdot T(1) + n \log n = n \log n.$$

We thus have solved the recurrence. In case you have doubts about the validity of the unfolding method, we verify by induction that indeed

$$T(n) \leq n \log n, \text{ for any integer } n \text{ that is a power of 2.}$$

The base case is when $n = 1$. In this case, we have $T(1) = 0$ and $1 \log 1 = 1 \cdot 0 = 0$. Let $n \geq 2$ and assume that

$$T(n/2) \leq (n/2) \log(n/2).$$

From the recurrence, we get

$$T(n) \leq 2 \cdot T(n/2) + n.$$

By substituting the induction hypothesis into this inequality, we get

$$\begin{aligned} T(n) &\leq 2 \cdot (n/2) \log(n/2) + n \\ &= n \log(n/2) + n \\ &= n (\log n - \log 2) + n \\ &= n (\log n - 1) + n \\ &= n \log n. \end{aligned}$$

Thus, by induction, $T(n) \leq n \log n$ for any integer n that is a power of 2.

Until now, we have only counted the number of comparisons made by algorithm MergeSort. It follows from the pseudocode that the total running time, i.e., the total number of "elementary" steps, is within a constant factor of the total number of comparisons. Therefore, if n is a power of 2, the running time of algorithm MergeSort(L, n) is $O(n \log n)$.

For general values of n, the recurrence for the number of comparisons becomes the following:

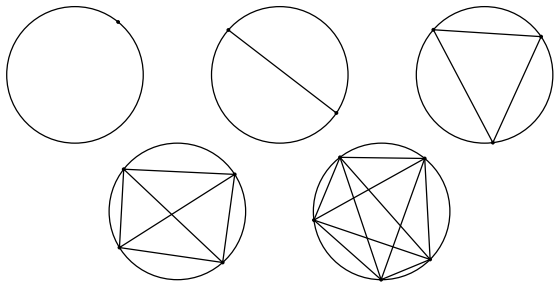
$$\begin{aligned} T(n) &= 0, \text{ if } n = 0 \text{ or } n = 1, \\ T(n) &\leq T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n, \text{ if } n \geq 2. \end{aligned}$$

It can be shown by induction that this recurrence solves to $T(n) = O(n \log n)$. We have proved the following result:

Theorem 4.5.1 *For any list L of n numbers, the running time of algorithm* MergeSort(L, n) *is $O(n \log n)$.*

4.6 Counting Regions when Cutting a Circle

Take a circle, place n points on it, and connect each pair of points by a straight-line segment. The points must be placed in such a way that no three segments pass through one point. These segments divide the circle into regions. Define R_n to be the number of such regions. Can we determine R_n?



By looking at the figure above, we see that

$$R_1 = 1, R_2 = 2, R_3 = 4, R_4 = 8, R_5 = 16.$$

There seems to be a clear pattern and it is natural to guess that R_n is equal to 2^{n-1} for all $n \geq 1$. To prove this, we have to argue that the number of regions doubles if we increase n by 1. If you try to do this, however, then you will fail! The reason is that R_n is *not* equal to 2^{n-1} for *all* $n \geq 1$; our guess was correct only for $1 \leq n \leq 5$.

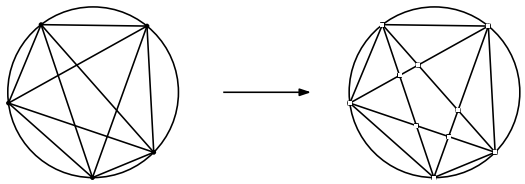
We will prove below that R_n grows only *polynomially* in n. This will imply that R_n cannot be equal to 2^{n-1} for all n, because the latter function grows *exponentially*.

4.6.1 A Polynomial Upper Bound on R_n

Let n be a (large) integer, consider a placement of n points on a circle, and connect each of the $\binom{n}{2}$ pairs of points by a straight-line segment. Recall that we assume that no three segments pass through one point. We define the following graph:

- Each of the n points on the circle is a vertex.
- Each intersection point between two segments is a vertex.
- These vertices divide the segments into subsegments and the circle into arcs in a natural way. Each such subsegment and arc is an edge of the graph.

The figure below illustrates this for the case when $n = 5$. The graph on the right has $10 = 5 + 5$ vertices: Each of the 5 points on the circle leads to one vertex and each of the 5 intersection points leads to one vertex. These 10 vertices divide the $\binom{5}{2} = 10$ segments into 20 straight-line edges and the circle into 5 circular edges. Therefore, the graph has $20 + 5 = 25$ edges.



Note that, strictly speaking, this process does not define a proper graph, because any two consecutive vertices on the circle are connected by two edges (one straight-line edge and one circular edge), whereas in a proper graph, there can be only one edge between any pair of vertices. For simplicity, however, we will refer to the resulting structure as a graph.

Define V_n and E_n to be the number of vertices and edges of the graph, respectively. We claim that

$$V_n \leq n + \binom{\binom{n}{2}}{2}. \tag{4.6}$$

This claim follows from the following observations:

- There are exactly n vertices on the circle.
- The n points on the circle are connected by $\binom{n}{2}$ segments, and any two such segments intersect at most once. Therefore, the number of vertices inside the circle is at most the number of pairs of segments. The latter quantity is equal to

$$\binom{\binom{n}{2}}{2}.$$

We next claim that

$$E_n \leq n + \binom{V_n}{2}. \tag{4.7}$$

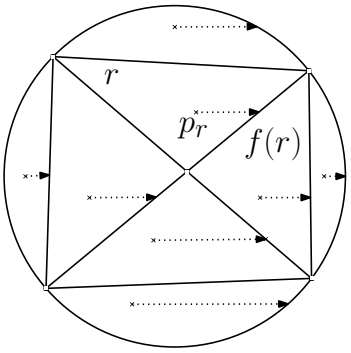
This claim follows from the following observations:

- There are exactly n edges on the circle.
- Any straight-line edge joins two vertices. Therefore, the number of straight-line edges is at most the number of pairs of vertices, which is $\binom{V_n}{2}$.

The final claim is that

$$R_n \leq E_n. \tag{4.8}$$

To prove this claim, we do the following. For each region r, choose a point p_r inside r, such that the y-coordinate of p_r is not equal to the y-coordinate of any vertex. Define $f(r)$ to be the edge that is reached when walking from p_r horizontally to the right.



This defines a one-to-one function f from the set of regions to the set of edges. Therefore, the number of regions, which is R_n, is at most the number of edges, which is E_n.

By combining (4.6), (4.7), and (4.8), we get

$$\begin{aligned} R_n &\leq E_n \\ &\leq n + \binom{V_n}{2} \\ &\leq n + \binom{n + \binom{\binom{n}{2}}{2}}{2}. \end{aligned}$$

In order to estimate the last quantity, we are going to use asymptotic notation; see Section 2.3. First observe that

$$\binom{n}{2} = \frac{n(n-1)}{2} = O(n^2).$$

This implies that

$$\binom{\binom{n}{2}}{2} = \binom{O(n^2)}{2} = O(n^4),$$

which implies that

$$n + \binom{\binom{n}{2}}{2} = n + O(n^4) = O(n^4),$$

which implies that

$$\binom{n + \binom{\binom{n}{2}}{2}}{2} = \binom{O(n^4)}{2} = O(n^8),$$

which implies that

$$R_n \leq n + \binom{n + \binom{\binom{n}{2}}{2}}{2} = n + O(n^8) = O(n^8).$$

Thus, we have proved our claim that R_n grows polynomially in n and, therefore, for large values of n, R_n is not equal to 2^{n-1}. (Using results that we will see in Section 7.4.1, it can be shown that, in fact, $R_n = O(n^4)$.)

We remark that there is a shorter way to prove that R_n is not equal to 2^{n-1} for all $n \geq 1$: You can verify by hand that $R_6 = 31$. Still, this single example does not rule out the possibility that R_n grows exponentially. The analysis that we gave above does rule this out.

We have proved above that $R_n = O(n^8)$. We also mentioned that this upper bound can be improved to $O(n^4)$. In the following subsections, we will prove that the latter upper bound cannot be improved. That is, we will prove that $R_n = \Theta(n^4)$. In fact, we will determine an exact formula, in terms of n, for the value of R_n.

4.6.2 A Recurrence Relation for R_n

Let $n \geq 2$ be an integer. Consider a placement of n points on a circle. We denote these points by $p_1, p_2, \ldots, p_n$ and assume that they are numbered in counterclockwise order. As before, we connect each of the $\binom{n}{2}$ pairs of points by a straight-line segment. We assume that no three segments pass through one point. We are going to derive a recurrence relation for the number R_n of regions in the following way:

- Remove all segments that have p_n as an endpoint. At this moment, the number of regions is, by definition, equal to R_{n-1}.

- Add the $n-1$ line segments $p_1p_n, p_2p_n, \ldots, p_{n-1}p_n$ one by one. For each segment p_kp_n added, determine the *increase* I_k in the number of regions.

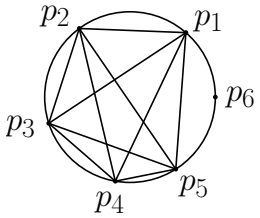
- Take the sum of R_{n-1} and all increases I_k, i.e.,

$$R_{n-1} + \sum_{k=1}^{n-1} I_k.$$

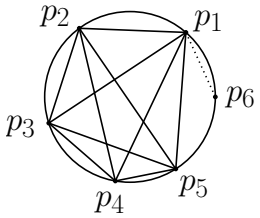
  This sum is equal to R_n, because in the entire process, we have counted each of the regions for n points exactly once.

- Thus, together with the base case $R_1 = 1$, we obtain a recurrence for the values R_n.

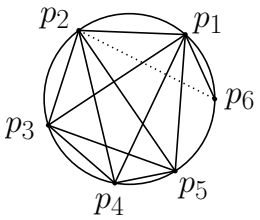
We start by illustrating this process for the case when $n = 6$. The figure below shows the situation after we have removed all segments that have p_6 as an endpoint. The number of regions is equal to $R_5 = 16$.



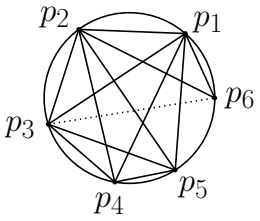
We are going to add, one by one, the five segments that have p_6 as an endpoint. When we add p_1p_6, one region gets cut into two. Thus, the number of regions increases by one. Using the notation introduced above, we have $I_1 = 1$.



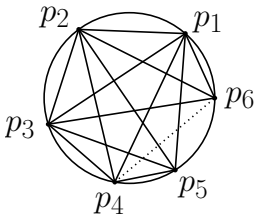
When we add p_2p_6, four regions get cut into two. Thus, the number of regions increases by four, and we have $I_2 = 4$.



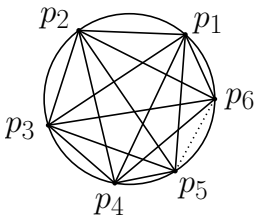
When we add p_3p_6, five regions get cut into two. Thus, the number of regions increases by five, and we have $I_3 = 5$.



When we add p_4p_6, four regions get cut into two. Thus, the number of regions increases by four, and we have $I_4 = 4$.



Finally, when we add p_5p_6, one region gets cut into two. Thus, the number of regions increases by one, and we have $I_5 = 1$.



After having added the five segments with endpoint p_6, we have accounted for all regions determined by the six points. In other words, the number of regions we have at the end is equal to R_6. Since the number of regions at

the end is equal to the sum of the number of regions we started with, which is R_5, and the total increase, we have

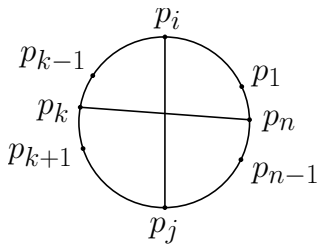
$$R_6 = R_5 + I_1 + I_2 + I_3 + I_4 + I_5 = 31.$$

Let us look at this more carefully. We have seen that $I_3 = 5$. That is, when adding the segment p_3p_6, the number of regions increases by 5. Where does this number 5 come from? The segment p_3p_6 *intersects* 4 segments, namely p_1p_4, p_1p_5, p_2p_4, and p_2p_5. The increase in the number of regions is one more than the number of intersections. Thus, when adding a segment, if we determine the number X of intersections between this new segment and existing segments, then the increase in the number of regions is equal to $1 + X$.

When we add p_3p_6, we have $X = 4$. Where does this number 4 come from? We make the following observations:

- Any segment that intersects p_3p_6 has one endpoint above p_3p_6 and one endpoint below p_3p_6.
- Any pair (a, b) of points, with a above p_3p_6 and b below p_3p_6, defines a segment ab that intersects p_3p_6.
- Thus, the value of X is equal to the number of pairs (a, b) of points in $\{p_1, p_2, p_3, p_4, p_5\}$, where a is above p_3p_6 and b is below p_3p_6. Since there are 2 choices for a (viz., p_1 and p_2) and 2 choices for b (viz., p_4 and p_5), it follows from the Product Rule that $X = 2 \cdot 2 = 4$.

Now that we have seen the basic approach, we are going to derive the recurrence for R_n for an arbitrary integer $n \geq 2$. After having removed all segments that have p_n as an endpoint, we have R_{n-1} regions. For each integer k with $1 \leq k \leq n - 1$, we add the segment p_kp_n. What is the number of existing segments that are intersected by this new segment?



We observe that for $i < j$,

- $p_i p_j$ intersects $p_k p_n$ if and only if $1 \leq i \leq k-1$ and $k+1 \leq j \leq n-1$.

Since there are $k-1$ choices for i and $n-k-1$ choices for j, the Product Rule implies that the number of intersections due to $p_k p_n$ is equal to $(k-1)(n-k-1)$. Thus, the segment $p_k p_n$ goes through $1+(k-1)(n-k-1)$ regions, and each of them is cut into two. It follows that, when adding $p_k p_n$, the increase I_k in the number of regions is equal to

$$I_k = 1 + (k-1)(n-k-1).$$

We conclude that

$$\begin{aligned} R_n &= R_{n-1} + \sum_{k=1}^{n-1} I_k \\ &= R_{n-1} + \sum_{k=1}^{n-1} \left(1 + (k-1)(n-k-1)\right). \end{aligned}$$

In the summation on the right-hand side

- the term 1 occurs exactly $n-1$ times, and
- the term $(k-1)(n-k-1)$ is non-zero only if $2 \leq k \leq n-2$.

It follows that, for $n \geq 2$,

$$R_n = R_{n-1} + (n-1) + \sum_{k=2}^{n-2} (k-1)(n-k-1). \tag{4.9}$$

Thus, together with the base case

$$R_1 = 1, \tag{4.10}$$

we have determined the recurrence we were looking for.

4.6.3 Simplifying the Recurrence Relation

In this subsection, we will use a combinatorial proof (see Section 3.7) to show that the summation on the right-hand side of (4.9) satisfies

$$\sum_{k=2}^{n-2}(k-1)(n-k-1) = \binom{n-1}{3}. \tag{4.11}$$

Consider the set $S = \{1, 2, \ldots, n-1\}$. We know that the number of 3-element subsets of S is equal to $\binom{n-1}{3}$. As we will see below, the summation on the left-hand side of (4.11) counts exactly the same subsets.

We divide the 3-element subsets of S into groups based on their middle element. Observe that the middle element can be any of the values $2, 3, \ldots, n-2$. Thus, for any k with $2 \leq k \leq n-2$, the k-th group G_k consists of all 3-element subsets of S whose middle element is equal to k. Since the groups are pairwise disjoint, we have

$$\binom{n-1}{3} = \sum_{k=2}^{n-2} |G_k|.$$

What is the size of the k-th group G_k? Any 3-element subset in G_k consists of

- one element from $\{1, 2, \ldots, k-1\}$,
- the element k, and
- one element from $\{k+1, k+2, \ldots, n-1\}$.

It then follows from the Product Rule that

$$|G_k| = (k-1) \cdot 1 \cdot (n-k-1) = (k-1)(n-k-1).$$

Thus, we have proved the identity in (4.11), and the recurrence relation in (4.9) and (4.10) becomes

$$\begin{array}{rcl} R_1 & = & 1, \\ R_n & = & R_{n-1} + (n-1) + \binom{n-1}{3}, \text{ if } n \geq 2. \end{array} \tag{4.12}$$

4.6.4 Solving the Recurrence Relation

Now that we have a recurrence that looks reasonable, we are going to apply the unfolding technique of Section 4.5 to solve it: By repeatedly applying the recurrence in (4.12), we get

$$\begin{aligned} R_n &= (n-1) + \binom{n-1}{3} + R_{n-1} \\ &= (n-1) + (n-2) + \binom{n-1}{3} + \binom{n-2}{3} + R_{n-2} \\ &= (n-1) + (n-2) + (n-3) + \binom{n-1}{3} + \binom{n-2}{3} + \binom{n-3}{3} + R_{n-3}. \end{aligned}$$

By continuing, we get

$$\begin{aligned} R_n &= (n-1) + (n-2) + (n-3) + \cdots + 3 + 2 + 1 \\ &\quad + \binom{n-1}{3} + \binom{n-2}{3} + \binom{n-3}{3} + \cdots + \binom{3}{3} + \binom{2}{3} + \binom{1}{3} \\ &\quad + R_1. \end{aligned}$$

Since $\binom{2}{3} = \binom{1}{3} = 0$ and $R_1 = 1$, we get

$$\begin{aligned} R_n &= (n-1) + (n-2) + (n-3) + \cdots + 3 + 2 + 1 \\ &\quad + \binom{n-1}{3} + \binom{n-2}{3} + \binom{n-3}{3} + \cdots + \binom{3}{3} \\ &\quad + 1. \end{aligned}$$

Since, by Theorem 2.2.10, the first summation is equal to

$$1 + 2 + 3 + \cdots + (n-1) = n(n-1)/2 = \binom{n}{2},$$

we get

$$R_n = 1 + \binom{n}{2} + \sum_{k=3}^{n-1} \binom{k}{3}.$$

The final step is to simplify the summation on the right-hand side. We will use a combinatorial proof to show that

$$\sum_{k=3}^{n-1} \binom{k}{3} = \binom{n}{4}. \tag{4.13}$$

Consider all 4-element subsets of the set $S = \{1, 2, \ldots, n\}$. We know that there are $\binom{n}{4}$ many such subsets. We divide these subsets into groups based on their largest element. For any k with $3 \leq k \leq n-1$, the k-th group G_k consists of all 4-element subsets of S whose largest element is equal to $k+1$. It should be clear that

$$\binom{n}{4} = \sum_{k=3}^{n-1} |G_k|.$$

To determine the size of the group G_k, we observe that any 4-element subset in G_k consists of

- three elements from $\{1, 2, \ldots, k\}$ and
- the element $k+1$.

It then follows from the Product Rule that

$$|G_k| = \binom{k}{3} \cdot 1 = \binom{k}{3},$$

completing the proof of (4.13).

After (finally!) having solved and simplified our recurrence, we conclude that for any integer $n \geq 1$,

$$R_n = 1 + \binom{n}{2} + \binom{n}{4}.$$

In Exercise 4.31, you will see a shorter way to determine the exact value of R_n. We went for the long derivation, because it allowed us to illustrate, along the way, several techniques from previous sections.

4.7 Exercises

4.1 The function $f : \mathbb{N} \to \mathbb{Z}$ is recursively defined as follows:

$$\begin{aligned} f(0) &= 7, \\ f(n) &= f(n-1) + 6n - 3 \text{ if } n \geq 1. \end{aligned}$$

Prove that $f(n) = 3n^2 + 7$ for all integers $n \geq 0$.

4.2 The function $f : \mathbb{N} \to \mathbb{Z}$ is recursively defined as follows:

$$\begin{aligned} f(0) &= 3, \\ f(n) &= 2 \cdot f(n-1) - (f(n-1))^2 \text{ if } n \geq 1. \end{aligned}$$

Prove that $f(n) = 1 - 2^{2^n}$ for all integers $n \geq 1$. (2^{2^n} means 2 to the power of 2^n.)

4.3 Recall that $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$. The function $f : \mathbb{N}^3 \to \mathbb{N}$ is defined as follows:

- $f(k, n, 0) = k + n$ for $k \geq 0$ and $n \geq 0$,
- $f(k, 0, 1) = 0$ for $k \geq 0$,
- $f(k, 0, 2) = 1$ for $k \geq 0$,
- $f(k, 0, i) = k$ for $k \geq 0$ and $i \geq 3$,
- $f(k, n, i) = f(k, f(k, n-1, i), i-1)$ for $k \geq 0$, $i \geq 1$, and $n \geq 1$.

Determine $f(2, 3, 2)$.

4.4 In Section 4.2, we have defined the Fibonacci numbers $f_0, f_1, f_2, \ldots$ Prove that for each integer $n \geq 1$,

$$\sum_{i=1}^{n} f_{2i} = f_{2n+1} - 1$$

and

$$f_1^2 + f_2^2 + f_3^2 + \cdots + f_n^2 = f_n f_{n+1}.$$

4.5 In Section 4.2, we have defined the Fibonacci numbers $f_0, f_1, f_2, \ldots$ Prove that for each integer $n \geq 0$,

- f_{3n} is even,
- f_{3n+1} is odd,
- f_{3n+2} is odd,
- f_{4n} is a multiple of 3.

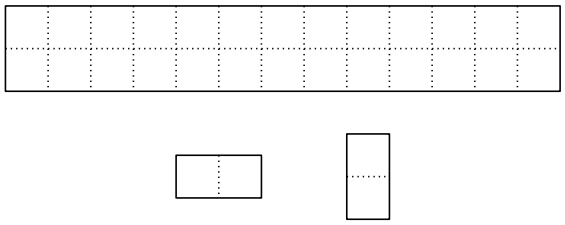
4.6 Let $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$, and let $n \geq 0$ be an integer. We have seen in Theorem 4.2.1 that

$$\frac{\varphi^n - \psi^n}{\sqrt{5}} \tag{4.14}$$

is equal to the n-th Fibonacci number f_n. Since the Fibonacci numbers are obviously integers, the number in (4.14) is an integer as well.

Prove that the number in (4.14) is a rational number using only Newton's Binomial Theorem.

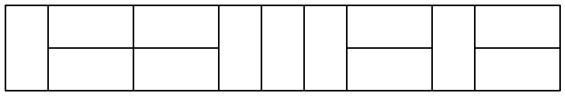
4.7 Let $n \geq 1$ be an integer and consider a $2 \times n$ board B_n consisting of $2n$ square cells. The top part of the figure below shows B_{13}.



A *brick* is a horizontal or vertical board consisting of 2 square cells; see the bottom part of the figure above. A *tiling* of the board B_n is a placement of bricks on the board such that

- the bricks exactly cover B_n and
- no two bricks overlap.

The figure below shows a tiling of B_{13}.



For $n \geq 1$, let a_n be the number of different tilings of the board B_n. Determine the value of a_n, i.e., express a_n in terms of numbers that we have seen in this chapter.

4.8 Consider strings of n characters, each character being a, b, c, or d, that contain an even number of as. (Recall that 0 is even.) Let E_n be the number of such strings. Prove that for any integer $n \geq 1$,

$$E_{n+1} = 2 \cdot E_n + 4^n.$$

4.9 Let A_n be the number of bitstrings of length n that contain 000. Prove that for $n \geq 4$,

$$A_n = A_{n-1} + A_{n-2} + A_{n-3} + 2^{n-3}.$$

4.10 Let $n \geq 1$ be an integer and define A_n to be the number of bitstrings of length n that do not contain 101.

- Determine A_1, A_2, A_3, and A_4.
- Prove that for each integer $n \geq 4$,

$$\begin{aligned} A_n &= 3 + A_1 + A_2 + A_3 + \cdots + A_{n-4} + A_{n-3} + A_{n-1} \\ &= 3 + \sum_{k=1}^{n-3} A_k + A_{n-1}. \end{aligned}$$

Hint: Divide the strings into groups depending on the number of leading 1s.

4.11 Let $n \geq 1$ be an integer and consider n people $P_1, P_2, \ldots, P_n$. Let A_n be the number of ways these n people can be divided into groups, such that each group consists of either one or two people.

- Determine A_1, A_2, and A_3.
- Prove that for each integer $n \geq 3$,

$$A_n = A_{n-1} + (n-1) \cdot A_{n-2}.$$

4.12 Let S be the set of ordered pairs of integers that is recursively defined in the following way:

- $(0, 0) \in S$.
- If $(a, b) \in S$ then $(a + 2, b + 3) \in S$.
- If $(a, b) \in S$ then $(a + 3, b + 2) \in S$.

Prove that for every element (a, b) in S, $a + b$ is divisible by 5.

4.13 Let S be the set of integers that is recursively defined in the following way:

- 4 is an element of S.
- If x and y are elements of S, then $x + y^2$ is an element of S.

Prove that every element of S is divisible by 4.

4.14 Let S be the set of ordered triples of integers that is recursively defined in the following way:

- $(66, 55, 1331) \in S$.
- If $(a, b, c) \in S$ then $(a + 7, b + 5, 14a - 10b + c + 24) \in S$.

Prove that for every element (a, b, c) in S,

$$a^2 - b^2 = c.$$

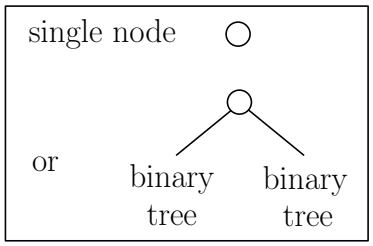
4.15 Let S be the set of integers that is recursively defined in the following way:

- 1 is an element of S.
- If x is an element of S, then $x + 2\sqrt{x} + 1$ is also an element of S.

Give a simple description of the set S and prove that your answer is correct.

4.16 A binary tree is

- either one single node
- or a node whose left subtree is a binary tree and whose right subtree is a binary tree.



Prove that any binary tree with n leaves has exactly $2n - 1$ nodes.

4.17 In Section 4.4, we have seen the recursive algorithm GOSSIP(n), which computes a sequence of phone calls for the persons $P_1, P_2, \ldots, P_n$. The base case for this algorithm was when $n = 4$. Assume we change the base case to $n = 2$: In this new base case, there are only two people P_1 and P_2, and only one phone call is needed. The rest of the algorithm remains unchanged.

Prove that the modified algorithm GOSSIP(n) results in a sequence of $2n - 3$ phone calls for any integer $n \geq 2$. (Thus, for $n \geq 4$, it makes one more phone call than the algorithm in Section 4.4.)

4.18 In Section 4.4, we have seen the recursive algorithm GOSSIP(n), which computes a sequence of phone calls for the persons $P_1, P_2, \ldots, P_n$, for any integer $n \geq 4$.

Give an iterative (i.e., non-recursive) version of this algorithm in pseudocode. Your algorithm must produce exactly the same sequence of phone calls as algorithm GOSSIP(n).

4.19 The following recursive algorithm FIB takes as input an integer $n \geq 0$ and returns the n-th Fibonacci number f_n:

Algorithm FIB(n)**:**

if $n = 0$ or $n = 1$
then $f = n$
else $f =$ FIB$(n-1)$ + FIB$(n-2)$
endif;
return f

Let a_n be the number of additions made by algorithm FIB(n), i.e., the total number of times the +-function in the else-case is called. Prove that for all $n \geq 0$,

$$a_n = f_{n+1} - 1.$$

4.20 Consider the following recursive algorithm BEER(n), which takes as input an integer $n \geq 1$:

```
Algorithm BEER(n):

    if n = 1
    then eat some peanuts
    else choose an arbitrary integer m with 1 ≤ m ≤ n − 1;
         BEER(m);
         drink one pint of beer;
         BEER(n − m)
    endif
```

- Explain why, for any integer $n \geq 1$, algorithm BEER(n) terminates.
- Let $B(n)$ be the number of pints of beer you drink when running algorithm BEER(n). Determine the exact value of $B(n)$.

4.21 Let $n \geq 2$ be an integer and consider a sequence $s_1, s_2, \ldots, s_n$ of n pairwise distinct numbers. The following algorithm computes the smallest and largest elements in this sequence:

```
Algorithm MINMAX(s_1, s_2, ..., s_n):

    min = s_1;
    max = s_1;
    for i = 2 to n
    do if s_i < min          (1)
       then min = s_i
       endif;
       if s_i > max          (2)
       then max = s_i
       endif
    endwhile;
    return (min, max)
```

This algorithm makes comparisons between input elements in lines (1) and (2). Determine the total number of comparisons as a function of n.

4.22 Let $n \geq 2$ be a power of 2 and consider a sequence S of n pairwise distinct numbers. The following algorithm computes the smallest and largest elements in this sequence:

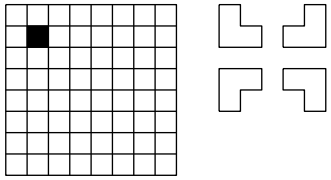
```
Algorithm FastMinMax(S, n):

    if n = 2
    then let x and y be the two elements in S;
         if x < y          (1)
         then min = x;
              max = y
         else min = y;
              max = x
         endif
    else divide S into two subsequences S1 and S2, both of size n/2;
         (min1, max1) = FastMinMax(S1, n/2);
         (min2, max2) = FastMinMax(S2, n/2);
         if min1 < min2          (2)
         then min = min1
         else min = min2
         endif;
         if max1 < max2          (3)
         then max = max2
         else max = max1
         endif
    endif;
    return (min, max)
```

This algorithm makes comparisons between input elements in lines (1), (2), and (3). Let $C(n)$ be the total number of comparisons made by algorithm FastMinMax on an input sequence of length n.

- Derive a recurrence for $C(n)$.
- Use this recurrence to prove that $C(n) = \frac{3}{2}\,n - 2$ for each $n \geq 2$ that is a power of 2.

4.23 Let k be a positive integer and let $n = 2^k$. You are given an $n \times n$ board B_n, all of whose (square) cells are white, except for one, which is black. (The left part of the figure below gives an example where $k = 3$ and $n = 8$.)

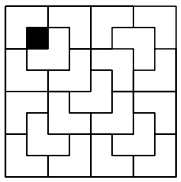
A *tromino* is an L-shaped object consisting of three 1×1 cells. Each tromino can appear in four different orientations; see the right part of the figure below.



A *tiling* of the board B_n is a placement of trominoes on the board such that

- the trominoes cover exactly all white cells (thus, the black cell is not covered by any tromino) and
- no two trominoes overlap.

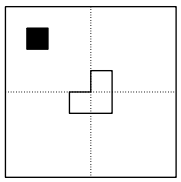
Here is a tiling of the board given above:



Describe a recursive algorithm that

- takes as input a board B_n having exactly one black cell (which can be anywhere on the board) and
- returns a tiling of this board.

Hint: Look at the following figure:

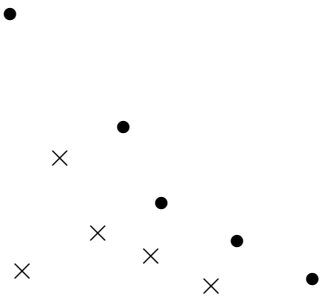


4.24 Let S be a set of n points in the plane. Each point p of S is given by its x- and y-coordinates p_x and p_y, respectively. We assume that no two points of S have the same x-coordinate and no two points of S have the same y-coordinate.

A point p of S is called *maximal* in S if there is no point in S that is to the north-east of p, i.e.,

$$\{q \in S : q_x > p_x \text{ and } q_y > p_y\} = \emptyset.$$

The figure below shows an example, in which the $\bullet$-points are maximal and the $\times$-points are not maximal. Observe that, in general, there is more than one maximal element in S.



Describe a recursive algorithm MaxElem(S) that has the same structure as algorithm MergeSort in Section 4.5 and does the following:

Input: A set S of n points in the plane, in sorted order from left to right.

Output: All maximal elements of S, in sorted order from left to right.

The running time $T(n)$ of your algorithm must be $O(n \log n)$.

4.25 The Hadamard matrices $H_0, H_1, H_2, \ldots$ are recursively defined as follows:

$$H_0 = (1)$$

and for $k \geq 1$,

$$H_k = \left(\begin{array}{c|c} H_{k-1} & H_{k-1} \\ \hline H_{k-1} & -H_{k-1} \end{array}\right).$$

Thus, H_0 is a 1×1 matrix whose only entry is 1,

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and

$$H_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Observe that H_k has 2^k rows and 2^k columns.

If x is a column vector of length 2^k, then $H_k x$ is the column vector of length 2^k obtained by multiplying the matrix H_k with the vector x.

Describe a recursive algorithm $\textsc{Mult}(k, x)$ that does the following:

Input: An integer $k \geq 0$ and a column vector x of length $n = 2^k$.

Output: The column vector $H_k x$ (having length n).

The running time $T(n)$ of your algorithm must be $O(n \log n)$.
Hint: The input only consists of k and x. The matrix H_k, which has n^2 entries, is not given as part of the input. Since you are aiming for an $O(n \log n)$–time algorithm, you cannot compute all entries of the matrix H_k.

4.26 Prove, for example by induction, that for $n \geq 1$,

$$1 + 2 + 3 + \cdots + n = n(n+1)/2,$$

and

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = n(n+1)(2n+1)/6.$$

4.27 Assume you remember that

$$1^2 + 2^2 + 3^2 + \cdots + n^2$$

is equal to a polynomial of degree three, i.e.,

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = An^3 + Bn^2 + Cn + D,$$

but you have forgotten the values of A, B, C, and D. How can you determine these four values?

4.28 In Section 4.6.3, we have shown that

$$\sum_{k=2}^{n-2} (k-1)(n-k-1) = \binom{n-1}{3}.$$

Use Exercise 4.26 to give an alternative proof.

4.29 In Section 4.6.4, we have used the fact that

$$\sum_{k=1}^{n-1} k = \binom{n}{2},$$

which follows from Theorem 2.2.10. Give an alternative proof that uses the approach that we used to prove the identity in (4.13).

4.30 In Section 4.6.4, we have shown that

$$\sum_{k=3}^{n-1} \binom{k}{3} = \binom{n}{4}.$$

Use induction and Pascal's Identity (see Theorem 3.7.2) to give an alternative proof.

4.31 Consider the numbers R_n that we defined in Section 4.6. The n points on the circle define $\binom{n}{2}$ line segments, one segment for each pair of points. Let X be the total number of intersections among these $\binom{n}{2}$ line segments.

- Prove that
$$R_n = 1 + \binom{n}{2} + X.$$
Hint: Start with only the circle and the n points. Then add the $\binom{n}{2}$ line segments one by one.

- Prove that
$$X = \binom{n}{4}.$$

4.32 For an integer $n \geq 1$, draw n straight lines, such that no two of them are parallel and no three of them intersect in one single point. These lines divide the plane into regions (some of which are bounded and some of which are unbounded). Denote the number of these regions by C_n.

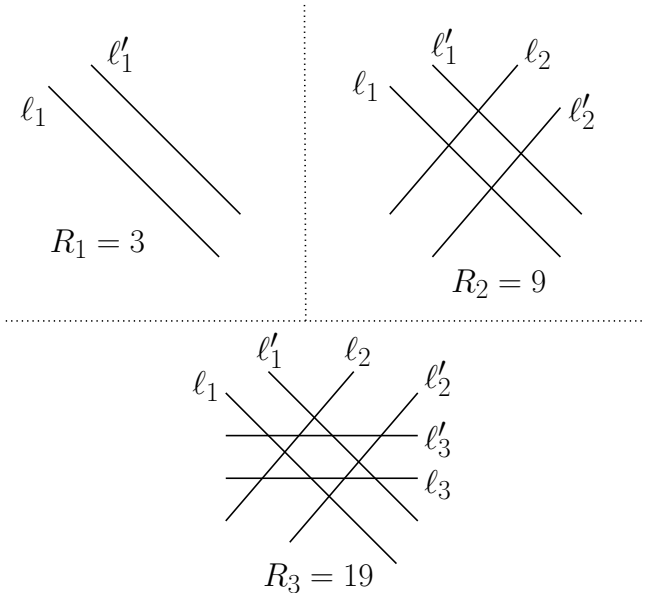
Derive a recurrence for the numbers C_n and use it to prove that for $n \geq 1$,

$$C_n = 1 + n(n+1)/2.$$

4.33 Let $n \geq 1$ be an integer. Consider $2n$ straight lines $\ell_1, \ell'_1, \ldots, \ell_n, \ell'_n$ such that

- for each i with $1 \leq i \leq n$, ℓ_i and ℓ'_i are parallel,
- no two of the lines $\ell_1, \ldots, \ell_n$ are parallel,
- no two of the lines $\ell'_1, \ldots, \ell'_n$ are parallel,
- no three of the $2n$ lines intersect in one single point.

These lines divide the plane into regions (some of which are bounded and some of which are unbounded). Denote the number of these regions by R_n. From the figure below, you can see that $R_1 = 3$, $R_2 = 9$, and $R_3 = 19$.



- Derive a recurrence for the numbers R_n and use it to prove that $R_n = 2n^2 + 1$ for $n \geq 1$.

Chapter 5

Discrete Probability

We all have some intuitive understanding of the notions of "chance" and "probability". When buying a lottery ticket, we know that there is a chance of winning the jackpot, but we also know that this chance is very small. Before leaving home in the morning, we check the weather forecast and see that, with probability 80%, we get 15 centimetres of snow in Ottawa. In this chapter, we will give a formal definition of this notion of "probability". We start by presenting a surprising application of probability and random numbers.

5.1 Anonymous Broadcasting

Consider a group of n people $P_1, P_2, \ldots, P_n$, for some integer $n \geq 3$. One person in this group, say P_k, would like to broadcast, *anonymously*, a message to all other people in the group. That is, P_k wants to broadcast a message such that

- everybody in the group receives the message,
- nobody knows that the message was broadcast by P_k.

In other words, when P_i (with $i \neq k$) receives the message, he only knows that it was broadcast by one of $P_1, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n$; P_i cannot determine who broadcast the message.

At first sight, it seems to be impossible to do this. In 1988, however, David Chaum published, in the Journal of Cryptology, a surprisingly simple

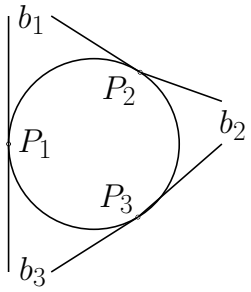
protocol that does achieve this. Chaum referred to the problem as the *Dining Cryptographers Problem*.

We will present and analyze the protocol for the case when $n = 3$. Thus, there are three people P_1, P_2, and P_3. We assume that exactly one of them broadcasts a message and refer to this person as the *broadcaster*. We also assume that the message is a binary string. The broadcaster will announce the message one bit at a time.

The three people P_1, P_2, and P_3 sit at a table, in clockwise order of their indices. Let b be the current bit that the broadcaster wants to announce. The protocol for broadcasting this bit is as follows:

Step 1: Each person P_i generates a *random* bit b_i, for example by flipping a coin. Thus, with 50% probability, $b_i = 0$ and with 50% probability, $b_i = 1$.

Step 2: Each person P_i looks at his own bit b_i and also shows it to his clockwise neighbor.



At the end of this second step,

- P_1 knows b_1 and b_3, but not b_2,
- P_2 knows b_1 and b_2, but not b_3,
- P_3 knows b_2 and b_3, but not b_1.

Step 3: Each person P_i computes the sum s_i (modulo 2) of the bits that he knows. Thus,

- P_1 computes $s_1 = (b_1 + b_3) \bmod 2$,
- P_2 computes $s_2 = (b_1 + b_2) \bmod 2$,
- P_3 computes $s_3 = (b_2 + b_3) \bmod 2$.

Step 4: Each person P_i does the following:

- If P_i is not the broadcaster, he sets $t_i = s_i$.
- If P_i is the broadcaster, he sets $t_i = (s_i + b) \bmod 2$. Recall that b is the current bit that the broadcaster wants to announce. (Thus, if $b = 1$, then P_i "secretly" flips the bit s_i and stores the result in t_i.)

Step 5: Each person P_i shows his bit t_i to the other two people.

Step 6: Each person P_i computes the sum (modulo 2) of the three bits t_1, t_2, and t_3, i.e., the value $(t_1 + t_2 + t_3) \bmod 2$.

This concludes the description of the protocol for broadcasting one bit b. Observe that for any bit x, we have $(x + x) \bmod 2 = 0$. Therefore, the bit computed in the last step is equal to

$$\begin{aligned} t_1 + t_2 + t_3 &= s_1 + s_2 + s_3 + b \\ &= (b_1 + b_3) + (b_1 + b_2) + (b_2 + b_3) + b \\ &= (b_1 + b_1) + (b_2 + b_2) + (b_3 + b_3) + b \\ &= b, \end{aligned}$$

where all arithmetic is done modulo 2. In other words, the bit computed in the last step is equal to the bit that the broadcaster wants to announce. This shows that each person in the group receives this bit.

It remains to show that a non-broadcaster cannot determine who broadcast the bit. In the analysis below, we assume that

- $b = 1$, i.e., the broadcaster announces the bit 1,
- P_2 is not the broadcaster.

We have to show that P_2 cannot determine whether P_1 or P_3 is the broadcaster. Note that P_2 knows the values

$$b_1, b_2, t_1, t_2, t_3,$$

but does not know the bit b_3. We consider the cases when $b_1 = b_2$ and $b_1 \neq b_2$ separately.

Case 1: $b_1 = b_2$. This case has two subcases depending on the value of b_3.

Case 1.1: $b_3 = b_1$; thus, all three b-bits are equal.

- If P_1 is the broadcaster, then (all arithmetic is done modulo 2)

$$t_1 = s_1 + 1 = b_1 + b_3 + 1 = 1$$

and

$$t_3 = s_3 = b_2 + b_3 = 0.$$

- If P_3 is the broadcaster, then

$$t_1 = s_1 = b_1 + b_3 = 0$$

and

$$t_3 = s_3 + 1 = b_2 + b_3 + 1 = 1.$$

Thus, the broadcaster is the person whose t-bit is equal to 1.

Case 1.2: $b_3 \neq b_1$ and, thus, $b_3 \neq b_2$.

- If P_1 is the broadcaster, then

$$t_1 = s_1 + 1 = b_1 + b_3 + 1 = 0$$

and

$$t_3 = s_3 = b_2 + b_3 = 1.$$

- If P_3 is the broadcaster, then

$$t_1 = s_1 = b_1 + b_3 = 1$$

and

$$t_3 = s_3 + 1 = b_2 + b_3 + 1 = 0.$$

Thus, the broadcaster is the person whose t-bit is equal to 0.

Since P_2 knows b_1 and b_2, he knows when Case 1 occurs. Since P_2 does not know b_3, however, he cannot determine whether Case 1.1 or 1.2 occurs. As a result, P_2 cannot determine whether P_1 or P_3 is the broadcaster.

Case 2: $b_1 \neq b_2$. This case has two subcases depending on the value of b_3.

Case 2.1: $b_3 = b_1$ and, thus, $b_3 \neq b_2$.

- If P_1 is the broadcaster, then

$$t_1 = s_1 + 1 = b_1 + b_3 + 1 = 1$$

and

$$t_3 = s_3 = b_2 + b_3 = 1.$$

- If P_3 is the broadcaster, then

$$t_1 = s_1 = b_1 + b_3 = 0$$

and

$$t_3 = s_3 + 1 = b_2 + b_3 + 1 = 0.$$

Thus, t_1 is always equal to t_3, no matter whether P_1 or P_3 is the broadcaster.

Case 2.2: $b_3 \neq b_1$ and, thus, $b_3 = b_2$.

- If P_1 is the broadcaster, then

$$t_1 = s_1 + 1 = b_1 + b_3 + 1 = 0$$

and

$$t_3 = s_3 = b_2 + b_3 = 0.$$

- If P_3 is the broadcaster, then

$$t_1 = s_1 = b_1 + b_3 = 1$$

and

$$t_3 = s_3 + 1 = b_2 + b_3 + 1 = 1.$$

Thus, t_1 is always equal to t_3, no matter whether P_1 or P_3 is the broadcaster.

Since P_2 knows b_1 and b_2, he knows when Case 2 occurs. Since P_2 does not know b_3, however, he cannot determine whether Case 2.1 or 2.2 occurs. As in Case 1, P_2 cannot determine whether P_1 or P_3 is the broadcaster.

We conclude from Cases 1 and 2 that the broadcasting of the bit $b = 1$ is indeed anonymous. Now consider the case when the bit b to be announced is equal to 0. It follows from the protocol that in this case, there is no "secret bit flipping" done in Step 4 and all three people use the same rules to compute the s-values and the t-values. In this case, $t_1 = s_1$, $t_2 = s_2$, and

$t_3 = s_3$, and P_2 can determine the bit b_3. He cannot, however, determine whether P_1 or P_3 is the broadcaster.

To conclude this section, we remark that for each bit to be announced, the entire protocol must be followed. That is, in each round of the protocol, one bit is broadcast and each person P_i must flip a coin to determine the bit b_i that is used in this round. We also remark that the protocol works only if exactly one person is the broadcaster.

5.2 Probability Spaces

In this section, we give a formal definition of the notion of "probability" in terms of sets and functions.

Definition 5.2.1 A *sample space* S is a non-empty countable set. Each element of S is called an *outcome* and each subset of S is called an *event*.

In daily life, we express probabilities in terms of percentages. For example, the weather forecast may tell us that, with 80% probability, we will be getting a snow storm today. In probability theory, probabilities are expressed in terms of numbers in the interval $[0, 1]$. A probability of 80% becomes a probability of 0.8.

Definition 5.2.2 Let S be a sample space. A *probability function* on S is a function $\Pr : S \to \mathbb{R}$ such that

- for all $\omega \in S$, $0 \leq \Pr(\omega) \leq 1$, and
- $\sum_{\omega \in S} \Pr(\omega) = 1$.

For any outcome ω in the sample space S, we will refer to $\Pr(\omega)$ as the probability that the outcome is equal to ω.

Definition 5.2.3 A *probability space* is a pair $(S, \Pr)$, where S is a sample space and Pr is a probability function on S.

A probability function Pr maps each element of the sample space S (i.e., each outcome) to a real number in the interval $[0, 1]$. It turns out to be useful

to extend this function so that it maps any event to a real number in $[0, 1]$. If A is an event (i.e., $A \subseteq S$), then we define

$$\Pr(A) = \sum_{\omega \in A} \Pr(\omega). \tag{5.1}$$

We will refer to $\Pr(A)$ as the probability that the event A occurs.

Note that since $S \subseteq S$, the entire sample space S is an event and

$$\Pr(S) = \sum_{\omega \in S} \Pr(\omega) = 1,$$

where the last equality follows from the second condition in Definition 5.2.2.

5.2.1 Examples

Flipping a coin: Assume we flip a coin. Since there are two possible outcomes (the coin either comes up *heads* (H) or *tails* (T)), the sample space is the set $S = \{H, T\}$. If the coin is *fair*, i.e., the probabilities of H and T are equal, then the probability function $\Pr : S \to \mathbb{R}$ is given by

$$\begin{aligned} \Pr(H) &= 1/2, \\ \Pr(T) &= 1/2. \end{aligned}$$

Observe that this function Pr satisfies the two conditions in Definition 5.2.2. Since this sample space has two elements, there are four events, one event for each subset. These events are

$$\emptyset, \{H\}, \{T\}, \{H, T\},$$

and it follows from (5.1) that

$$\begin{array}{rcccccc} \Pr(\emptyset) & = & 0, & & & & \\ \Pr(\{H\}) & = & \Pr(H) & = & 1/2, & & \\ \Pr(\{T\}) & = & \Pr(T) & = & 1/2, & & \\ \Pr(\{H, T\}) & = & \Pr(H) + \Pr(T) & = & 1/2 + 1/2 & = & 1. \end{array}$$

Flipping a coin twice: If we flip a fair coin twice, then there are four possible outcomes, and the sample space becomes $S = \{HH, HT, TH, TT\}$. For example, HT indicates that the first flip resulted in heads, whereas the

second flip resulted in tails. In this case, the probability function $\Pr : S \to \mathbb{R}$ is given by

$$\Pr(HH) = \Pr(HT) = \Pr(TH) = \Pr(TT) = 1/4.$$

Observe again that this function Pr satisfies the two conditions in Definition 5.2.2. Since the sample space consists of 4 elements, the number of events is equal to $2^4 = 16$. For example, $A = \{HT, TH\}$ is an event and it follows from (5.1) that

$$\Pr(A) = \Pr(HT) + \Pr(TH) = 1/4 + 1/4 = 1/2.$$

In words, when flipping a fair coin twice, the probability that we see one heads and one tails (without specifying the order) is equal to 1/2.

Rolling a die twice: If we roll a fair die, then there are six possible outcomes (1, 2, 3, 4, 5, and 6), each one occurring with probability 1/6. If we roll this die twice, we obtain the sample space

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\},$$

where i is the result of the first roll and j is the result of the second roll. Note that $|S| = 6 \times 6 = 36$. Since the die is fair, each outcome has the same probability. Therefore, in order to satisfy the two conditions in Definition 5.2.2, we must have

$$\Pr(i, j) = 1/36$$

for each outcome (i, j) in S.

If we are interested in the sum of the results of the two rolls, then we define the event

$$A_k = \text{ “the sum of the results of the two rolls is equal to } k\text{”},$$

which, using the notation of sets, is the same as

$$A_k = \{(i, j) \in S : i + j = k\}.$$

In the matrix below, the leftmost column indicates the result of the first roll, the top row indicates the result of the second roll, and each entry is the sum of the results of the two corresponding rolls.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

As can be seen from this matrix, the event A_k is non-empty only if $k \in \{2, 3, \ldots, 12\}$. For any other k, the event A_k is empty, which means that it can never occur.

It follows from (5.1) that

$$\Pr(A_k) = \sum_{(i,j)\in A_k} \Pr(i,j) = \sum_{(i,j)\in A_k} 1/36 = |A_k|/36.$$

For example, if $k = 4$, you can see in the matrix that the event A_4 has three elements, i.e., there are 3 possible outcomes of two rolls that result in a sum of 4. These outcomes are $(3,1)$, $(2,2)$, and $(1,3)$. It follows that

$$\Pr(A_4) = 3/36 = 1/12.$$

In a similar way, we see that

$$\begin{aligned}
\Pr(A_2) &= 1/36, \\
\Pr(A_3) &= 2/36 = 1/18, \\
\Pr(A_4) &= 3/36 = 1/12, \\
\Pr(A_5) &= 4/36 = 1/9, \\
\Pr(A_6) &= 5/36, \\
\Pr(A_7) &= 6/36 = 1/6, \\
\Pr(A_8) &= 5/36, \\
\Pr(A_9) &= 4/36 = 1/9, \\
\Pr(A_{10}) &= 3/36 = 1/12, \\
\Pr(A_{11}) &= 2/36 = 1/18, \\
\Pr(A_{12}) &= 1/36.
\end{aligned}$$

A sample space is not necessarily uniquely defined. In the last example, where we are interested in the sum of the results of two rolls of a die, we could also have taken the sample space to be the set

$$S' = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

The probability function $\Pr'$ corresponding to this sample space S' is given by

$$\Pr'(k) = \Pr(A_k),$$

because $\Pr'(k)$ is the probability that we get the outcome k in the sample space S', which is the same as the probability that event A_k occurs in the sample space S. You should verify that this function $\Pr'$ satisfies the two conditions in Definition 5.2.2 and, thus, is a valid probability function on S'.

5.3 Basic Rules of Probability

In this section, we prove some basic properties of probability functions. As we will see, all these properties follow from Definition 5.2.2. Throughout this section, $(S, \Pr)$ is a probability space.

Recall that an event is a subset of the sample space S. In particular, the emptyset $\emptyset$ is an event. Intuitively, $\Pr(\emptyset)$ must be zero, because it is the probability that there is no outcome, which can never occur. The following lemma states that this is indeed the case.

Lemma 5.3.1 $\Pr(\emptyset) = 0$.

Proof. By (5.1), we have

$$\Pr(\emptyset) = \sum_{\omega \in \emptyset} \Pr(\omega).$$

Since there are zero terms in this summation, its value is equal to zero. ∎

We say that two events A and B are *disjoint*, if $A \cap B = \emptyset$. A sequence $A_1, A_2, \ldots, A_n$ of events is *pairwise disjoint*, if any pair in this sequence is disjoint. The following lemma is similar to the Sum Rule of Section 3.4.

Lemma 5.3.2 *If $A_1, A_2, \ldots, A_n$ is a sequence of pairwise disjoint events, then*

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i).$$

Proof. Define $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Using (5.1), we get

$$\begin{aligned} \Pr(A) &= \sum_{\omega \in A} \Pr(\omega) \\ &= \sum_{i=1}^{n} \sum_{\omega \in A_i} \Pr(\omega) \\ &= \sum_{i=1}^{n} \Pr\left(A_i\right). \end{aligned}$$

■

To give an example, assume we roll a fair die twice. What is the probability that the sum of the two results is even? If you look at the matrix in Section 5.2, then you see that there are 18 entries, out of 36, that are even. Therefore, the probability of having an even sum is equal to $18/36 = 1/2$. Below we will give a different way to determine this probability.

The sample space is the set

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\},$$

where i is the result of the first roll and j is the result of the second roll. Each element of S has the same probability $1/36$ of being an outcome of rolling the die twice.

The event we are interested in is

$$A = \{(i, j) \in S : \ i + j \text{ is even}\}.$$

Observe that $i + j$ is even if and only if both i and j are even or both i and j are odd. Therefore, we split the event A into two disjoint events

$$A_1 = \{(i, j) \in S : \ \text{both } i \text{ and } j \text{ are even}\}$$

and

$$A_2 = \{(i, j) \in S : \ \text{both } i \text{ and } j \text{ are odd}\}.$$

By Lemma 5.3.2, we have

$$\Pr(A) = \Pr\left(A_1\right) + \Pr\left(A_2\right).$$

The set A_1 has $3 \cdot 3 = 9$ elements, because there are 3 choices for i and 3 choices for j. Similarly, the set A_2 has 9 elements. It follows that

$$\Pr(A) = \Pr(A_1) + \Pr(A_2) = 9/36 + 9/36 = 1/2.$$

In the next lemma, we relate the probability that an event occurs to the probability that the event does not occur. If A is an event, then $\overline{A}$ denotes its *complement*, i.e., $\overline{A} = S \setminus A$. Intuitively, the sum of $\Pr(A)$ and $\Pr\left(\overline{A}\right)$ must be equal to one, because the event A either occurs or does not occur. The following lemma states that this is indeed the case. Observe that this is similar to the Complement Rule of Section 3.3.

Lemma 5.3.3 *For any event A,*

$$\Pr(A) = 1 - \Pr\left(\overline{A}\right).$$

Proof. Since A and $\overline{A}$ are disjoint and $S = A \cup \overline{A}$, it follows from Lemma 5.3.2 that

$$\Pr(S) = \Pr\left(A \cup \overline{A}\right) = \Pr(A) + \Pr\left(\overline{A}\right).$$

We have seen in Section 5.2 that $\Pr(S) = 1$. ∎

Consider again the sample space that we saw after Lemma 5.3.2. We showed that, when rolling a fair die twice, we get an even sum with probability 1/2. It follows from Lemma 5.3.3 that we get an odd sum with probability $1 - 1/2 = 1/2$.

The next lemma is similar to the Principle of Inclusion and Exclusion that we have seen in Section 3.5.

Lemma 5.3.4 *If A and B are events, then*

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Proof. Since $B \setminus A$ and $A \cap B$ are disjoint and $B = (B \setminus A) \cup (A \cap B)$, it follows from Lemma 5.3.2 that

$$\Pr(B) = \Pr(B \setminus A) + \Pr(A \cap B).$$

Next observe that A and $B \setminus A$ are disjoint. Since $A \cup B = A \cup (B \setminus A)$, we again apply Lemma 5.3.2 and obtain

$$\Pr(A \cup B) = \Pr(A) + \Pr(B \setminus A).$$

By combining these two equations, we obtain the claim in the lemma. ∎

To give an example, assume we choose a number x in the sample space $S = \{1, 2, \ldots, 1000\}$, such that each element has the same probability $1/1000$ of being chosen. What is the probability that x is divisible by 2 or 3? Define the events

$$A = \{i \in S : \ i \text{ is divisible by } 2\}$$

and

$$B = \{i \in S : \ i \text{ is divisible by } 3\}.$$

Then we want to determine $\Pr(A \cup B)$, which, by Lemma 5.3.4, is equal to

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

Since there are $\lfloor 1000/2 \rfloor = 500$ even numbers in S, we have

$$\Pr(A) = 500/1000.$$

Since there are $\lfloor 1000/3 \rfloor = 333$ elements in S that are divisible by 3, we have

$$\Pr(B) = 333/1000.$$

Observe that i belongs to $A \cap B$ if and only if i is divisible by 6, i.e.,

$$A \cap B = \{i \in S : \ i \text{ is divisible by } 6\}.$$

Since there are $\lfloor 1000/6 \rfloor = 166$ elements in S that are divisible by 6, we have

$$\Pr(A \cap B) = 166/1000.$$

We conclude that

$$\begin{aligned} \Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\ &= 500/1000 + 333/1000 - 166/1000 \\ &= 667/1000. \end{aligned}$$

Lemma 5.3.5 (Union Bound) *For any integer $n \geq 1$, if $A_1, A_2, \ldots, A_n$ is a sequence of events, then*

$$\Pr\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) \leq \sum_{i=1}^{n} \Pr\left(A_i\right).$$

Proof. The proof is by induction on n. If $n = 1$, we have equality and, thus, the claim obviously holds. Let $n \geq 2$ and assume the claim is true for $n-1$, i.e., assume that

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \leq \sum_{i=1}^{n-1} \Pr(A_i).$$

Define $B = A_1 \cup A_2 \cup \cdots \cup A_{n-1}$. Then it follows from Lemma 5.3.4 that

$$\Pr(B \cup A_n) = \Pr(B) + \Pr(A_n) - \Pr(B \cap A_n) \leq \Pr(B) + \Pr(A_n),$$

because $\Pr(B \cap A_n) \geq 0$ (this follows from the first condition in Definition 5.2.2). Since we assumed that

$$\Pr(B) \leq \sum_{i=1}^{n-1} \Pr(A_i),$$

it follows that

$$\begin{aligned} \Pr(A_1 \cup A_2 \cup \cdots \cup A_n) &= \Pr(B \cup A_n) \\ &\leq \Pr(B) + \Pr(A_n) \\ &\leq \sum_{i=1}^{n-1} \Pr(A_i) + \Pr(A_n) \\ &= \sum_{i=1}^{n} \Pr(A_i). \end{aligned}$$

■

Lemma 5.3.6 *If A and B are events with $A \subseteq B$, then*

$$\Pr(A) \leq \Pr(B).$$

Proof. Using (5.1) and the fact that $\Pr(\omega) \geq 0$ for each ω in S, we get

$$\begin{aligned} \Pr(A) &= \sum_{\omega \in A} \Pr(\omega) \\ &\leq \sum_{\omega \in B} \Pr(\omega) \\ &= \Pr(B). \end{aligned}$$

■

5.4 Uniform Probability Spaces

In this section, we consider finite sample spaces S in which each outcome has the same probability. Since, by Definition 5.2.2, all probabilities add up to one, the probability of each outcome must be equal to $1/|S|$.

Definition 5.4.1 A *uniform probability space* is a pair $(S, \Pr)$, where S is a finite sample space and the probability function Pr satisfies

$$\Pr(\omega) = \frac{1}{|S|} \text{ for each outcome } \omega \text{ in } S.$$

The probability spaces that we have seen in Section 5.2.1 are all uniform, except the space $(S', \Pr')$ that we saw at the end of that section.

To give another example, when playing[1] Lotto 6/49, you choose a 6-element subset of the set $A = \{1, 2, \ldots, 49\}$. Twice a week, the Ontario Lottery and Gaming Corporation (OLG) draws the six "winning numbers" uniformly at random from A. If your numbers are equal to the ones drawn by OLG, then you can withdraw from this course and spend the rest of your life on the beach. Most people would find it silly to choose the subset $\{1, 2, 3, 4, 5, 6\}$; they would argue that it is better to choose, for example, the subset $\{2, 5, 16, 36, 41, 43\}$. Is this true?

For this example, the sample space is the set S consisting of all 6-element subsets of A. Since S has size $\binom{49}{6}$ and the subset drawn by OLG is uniform, each outcome (i.e., each 6-element subset of S) has a probability of

$$\frac{1}{\binom{49}{6}} = \frac{1}{13,983,816} \approx 0.000000072.$$

In particular, both $\{1, 2, 3, 4, 5, 6\}$ and $\{2, 5, 16, 36, 41, 43\}$ have the *same* probability of being the winning numbers. (Still, the latter subset was drawn by OLG on February 8, 2014.)

The lemma below states that in a uniform probability space $(S, \Pr)$, the probability of an event A is just the ratio of the size of A and the size of S.

Lemma 5.4.2 *If $(S, \Pr)$ is a uniform probability space and A is an event, then*

$$\Pr(A) = \frac{|A|}{|S|}.$$

[1] Of course, you should *never* waste money on this!

Proof. By applying (5.1), we get

$$\Pr(A) = \sum_{\omega \in A} \Pr(\omega) = \sum_{\omega \in A} \frac{1}{|S|} = \frac{|A|}{|S|}.$$

∎

5.4.1 The Probability of Getting a Full House

In a standard deck of 52 cards, each card has a *suit* and a *rank*. There are four suits (spades ♠, hearts ♡, clubs ♣, and diamonds ♢), and 13 ranks (ace, $2, 3, 4, 5, 6, 7, 8, 9, 10$, jack, queen, and king).

A hand of five cards is called a *full house*, if three of the cards are of the same rank and the other two cards are also of the same (but necessarily different) rank. For example, a hand consisting of three sevens and two queens is a full house.

Assume we get a random hand of five cards. What is the probability that this hand is a full house? To answer this question, we start by observing that the sample space is the set S consisting of all hands of five cards. Note that

$$|S| = \binom{52}{5} = 2,598,960.$$

We assume that the 5-card hand is chosen uniformly at random, so that each hand has a probability of $1/|S|$ of being chosen.

Since we are interested in the probability of a random hand being a full house, we define the event A to be the set of all elements in S that are full houses. By Lemma 5.4.2, we have

$$\Pr(A) = \frac{|A|}{|S|}.$$

Thus, to determine $\Pr(A)$, it remains to determine the size of the set A, i.e., the total number of full houses. For this, we will use the Product Rule of Section 3.1:

- The procedure is "choose a full house".
- First task: Choose the rank of the three cards in the full house. There are 13 ways to do this.

- Second task: Choose the suits of these three cards. There are $\binom{4}{3}$ ways to do this.
- Third task: Choose the rank of the other two cards in the full house. There are 12 ways to do this.
- Fourth task: Choose the suits of these two cards. There are $\binom{4}{2}$ ways to do this.

Thus, the number of full houses is equal to

$$|A| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} = 3,744.$$

We conclude that the probability of getting a full house is equal to

$$\Pr(A) = \frac{|A|}{|S|} = \frac{3,744}{2,598,960} \approx 0.00144.$$

5.5 The Birthday Paradox

Let $n \geq 2$ be an integer and consider a group of n people. In this section, we will determine the probability p_n that at least two of them have the same birthday. We will ignore leap years, so that there are 365 days in one year.

Below, we will show that $p_2 = 1/365$. If $n \geq 366$, then it follows from the Pigeonhole Principle (see Section 3.10) that there must be at least two people with the same birthday and, therefore, $p_n = 1$. Intuitively, if n increases from 2 to 365, the value of p_n should increase as well. What is the value of n such that p_n is larger than $1/2$ for the first time? That is, what is the value of n for which $p_{n-1} \leq 1/2 < p_n$? In this section, we will see that this question can be answered using simple counting techniques that we have seen in Chapter 3.

We denote the people by $P_1, P_2, \ldots, P_n$, we denote the number of days in one year by d, and we number the days in one year as $1, 2, \ldots, d$. The sample space is the set

$$S_n = \{(b_1, b_2, \ldots, b_n) : b_i \in \{1, 2, \ldots, d\} \text{ for each } 1 \leq i \leq n\},$$

where b_i denotes the birthday of P_i. Note that

$$|S_n| = d^n.$$

We consider the uniform probability space; thus, for each element $(b_1, b_2, \ldots, b_n)$ in S_n, we have

$$\Pr(b_1, b_2, \ldots, b_n) = \frac{1}{|S_n|} = \frac{1}{d^n}.$$

The event we are interested in is

$$A_n = \text{ “at least two of the numbers in } b_1, b_2, \ldots, b_n \text{ are equal”}.$$

Using the notation of sets, this is the same as

$$A_n = \{(b_1, b_2, \ldots, b_n) \in S_n : \ b_1, b_2, \ldots, b_n \text{ contains duplicates}\}.$$

The probability p_n that we introduced above is equal to

$$p_n = \Pr(A_n).$$

As mentioned above, the Pigeonhole Principle implies that $p_n = 1$ for $n > d$. Therefore, we assume from now on that $n \leq d$.

Let us start by determining p_2. Since we consider the uniform probability space, we have, by Lemma 5.4.2,

$$p_2 = \Pr(A_2) = \frac{|A_2|}{|S_2|}.$$

We know already that $|S_2| = d^2$. The event A_2 is equal to

$$A_2 = \{(1,1), (2,2), \ldots, (d,d)\}.$$

Thus, $|A_2| = d$ and we obtain

$$p_2 = \frac{|A_2|}{|S_2|} = \frac{d}{d^2} = \frac{1}{d}.$$

To determine p_n for larger values of n, it is easier to determine the probability of the complement $\overline{A}_n$. The latter probability, together with Lemma 5.3.3, will give us the value of p_n. Note that

$$\overline{A}_n = \{(b_1, b_2, \ldots, b_n) \in S_n : \ b_1, b_2, \ldots, b_n \text{ are pairwise distinct}\}.$$

In other words, $\overline{A}_n$ is the set of all ordered sequences consisting of n pairwise distinct elements of $\{1, 2, \ldots, d\}$. In Section 3.6, see (3.1), we have seen that

$$|\overline{A}_n| = \frac{d!}{(d-n)!}.$$

We conclude that, for any n with $2 \leq n \leq d$,

$$\begin{aligned} p_n &= \Pr(A_n) \\ &= 1 - \Pr\left(\overline{A}_n\right) \\ &= 1 - \frac{|\overline{A}_n|}{|S_n|} \\ &= 1 - \frac{d!}{(d-n)!d^n}. \end{aligned}$$

By taking $d = 365$, we get $p_{22} = 0.476$ and $p_{23} = 0.507$. Thus, in a random group of 23 people[2], the probability that at least two of them have the same birthday is more than 50%. Most people are very surprised when they see this for the first time, because our intuition says that a much larger group is needed to have a probability of more than 50%. The values p_n approach 1 pretty fast. For example, $p_{40} = 0.891$ and $p_{100} = 0.9999997$.

5.5.1 Throwing Balls into Boxes

When we derived the expression for p_n, we did not use the fact that the value of d is equal to 365. In other words, the expression is valid for any value of d. For general values of d, we can interpret the birthday problem in the following way: Consider d boxes $B_1, B_2, \ldots, B_d$, where d is a large integer. Assume that we throw n balls into these boxes so that each ball lands in a random box. Then p_n is the probability that there is at least one box that contains more than one ball. Since it is not easy to see how the expression

$$p_n = 1 - \frac{d!}{(d-n)!d^n}$$

depends on n, we will approximate it by a simpler expression. For this, we will use the inequality

$$1 - x \leq e^{-x}, \tag{5.2}$$

which is valid for any real number x. If x is close to zero, then the inequality is very accurate. The easiest way to prove this inequality is by showing that the minimum of the function $f(x) = x + e^{-x}$ is equal to $f(0) = 1$, using techniques from calculus.

[2] two soccer teams plus the referee

If we define $q_n = 1 - p_n$, then we have

$$q_n = \frac{d!}{(d-n)!d^n}.$$

Using (5.2), we get

$$\begin{aligned} q_n &= \frac{d}{d} \cdot \frac{d-1}{d} \cdot \frac{d-2}{d} \cdot \frac{d-3}{d} \cdots \frac{d-(n-1)}{d} \\ &= \frac{d-1}{d} \cdot \frac{d-2}{d} \cdot \frac{d-3}{d} \cdots \frac{d-(n-1)}{d} \\ &= (1-1/d) \cdot (1-2/d) \cdot (1-3/d) \cdots (1-(n-1)/d) \\ &\leq e^{-1/d} \cdot e^{-2/d} \cdot e^{-3/d} \cdots e^{-(n-1)/d} \\ &= e^{-(1+2+3+\cdots+(n-1))/d}. \end{aligned}$$

Using the equality

$$1 + 2 + 3 + \cdots + (n-1) = n(n-1)/2,$$

see Theorem 2.2.10, we thus get

$$q_n \leq e^{-n(n-1)/(2d)},$$

and therefore,

$$p_n = 1 - q_n \geq 1 - e^{-n(n-1)/(2d)}.$$

If n is large, then $n(n-1)/(2d)$ is very close to $n^2/(2d)$ and, thus,

$$p_n \gtrsim 1 - e^{-n^2/(2d)}.$$

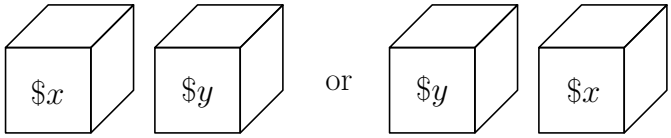
If we take $n = \sqrt{2d}$, then we get

$$p_n \gtrsim 1 - e^{-1} \approx 0.632.$$

Thus, for large values of d, if we throw $\sqrt{2d}$ balls into d boxes, then with probability (approximately) at least $1 - 1/e$, there is a box that contains more than one ball.

5.6 The Big Box Problem

Keith chooses two distinct elements x and y, with $x < y$, from the set $A = \{0, 1, 2, \ldots, 100\}$. He takes two identical boxes, and puts x dollars in one box and y dollars in the other box. Then Keith closes the two boxes, shuffles them, and puts them on a table. At this moment, we neither know x nor y; the only information we have is that these are distinct elements from the set A.



We will refer to the box containing x dollars as the *small box* and to the box containing y dollars as the *big box*. Our goal is to find the big box. We are allowed to do the following:

1. We can choose one of the two boxes, open it, and count how much money is inside it.

2. Now we have to make our final decision: Either we keep the box we just opened or we take the other box.

For example, assume that the box we pick in the first step contains \$33. Then we know that the other box contains either less than \$33 or more than \$33. It seems that the only reasonable thing to do is to flip a fair coin when making our final decision. If we do that, then we find the big box with probability 0.5.

In the rest of this section, we will show the surprising result that we can find the big box with probability at least 0.505.

The idea is as follows. *Assume* that we know a number z such that $x < z < y$. (Keep in mind that we do not know x and we do not know y. Thus, we assume that we know a number z that is between the two unknown numbers x and y.)

- If the box we choose in the first step contains more than z dollars, then we know that this is the big box and, therefore, we keep it.

- If the box we choose in the first step contains less than z dollars, then we know that this is the small box and, therefore, we take the other box.

Thus, if we know this number z with $x < z < y$, then we are guaranteed to find the big box.

Of course, it is not realistic to assume that we know this magic number z. The trick is to choose a *random* z and *hope* that it is between x and y. If z is between x and y, then we find the big box with probability 1; otherwise, we find the big box with probability $1/2$. As we will see later, the overall probability of finding the big box will be at least 0.505.

In order to avoid the case when $z = x$ or $z = y$, we will choose z from the set

$$B = \{1/2, 3/2, 5/2, \ldots, 100 - 1/2\}.$$

Note that $|B| = 100$. Our algorithm that attempts to find the big box does the following:

Algorithm FindBigBox**:**

Step 1: Choose one of the two boxes uniformly at random, open it, and count the amount of money inside it; let this amount be a.

Step 2: Choose z uniformly at random from the set B.

Step 3: Do the following:

- If $a > z$, then keep the box chosen in Step 1.
- Otherwise (i.e., if $a < z$), take the other box.

5.6.1 The Probability of Finding the Big Box

We are now going to determine the probability that this algorithm finds the big box. First, we have to ask ourselves what the sample space is. There are two places in the algorithm where a random element is obtained:

- In Step 1, we obtain the element a, which is a random element from the set $\{x, y\}$. We know that this value a is equal to one of x and y. However, at the end of Step 1, we do not know whether $a = x$ or $a = y$.
- In Step 2, we obtain a random element from the set B.

Based on this, the sample space S is the Cartesian product

$$S = \{x, y\} \times B = \{(a, z) : a \in \{x, y\}, z \in B\}$$

and Steps 1 and 2 can be replaced by

- choose a uniformly random element (a, z) in S.

Note that $|S| = 200$.

We say that algorithm FindBigBox is *successful* if it finds the big box. Thus, we want to determine $\Pr(W)$, where W is the event

$$W = \text{ “algorithm FindBigBox is successful”.}$$

We are going to write this event as a subset of the sample space S. For this, we have to determine all elements (a, z) in S for which algorithm FindBigBox is successful.

First consider the case when $a = x$. In this case, the box we choose in Step 1 is the small box. There are two possibilities for z:

- If $x = a > z$, then the algorithm keeps the small box and, thus, is not successful.
- If $x = a < z$, then the algorithm takes the other box (which is the big box) and, thus, is successful.

Thus, the event W contains the set

$$W_x = \{(x, z) : z \in \{x + 1/2, x + 3/2, \ldots, 100 - 1/2\}.$$

You can verify that

$$|W_x| = 100 - x.$$

The second case to consider is when $a = y$. In this case, the box we choose in Step 2 is the big box. Again, there are two possibilities for z:

- If $y = a > z$, then the algorithm keeps the big box and, thus, is successful.
- If $y = a < z$, then the algorithm takes the other box (which is the small box) and, thus, is not successful.

Thus, the event W contains the set

$$W_y = \{(y, z) : z \in \{1/2, 3/2, \ldots, y - 1/2\}.$$

You can verify that

$$|W_y| = y.$$

We conclude that

$$W = W_x \cup W_y.$$

Since the events W_x and W_y are disjoint, the probability that algorithm FindBigBox is successful is equal to

$$\begin{aligned} \Pr(W) &= \Pr(W_x) + \Pr(W_y) \\ &= \frac{|W_x|}{|S|} + \frac{|W_y|}{|S|} \\ &= \frac{100 - x}{200} + \frac{y}{200} \\ &= \frac{1}{2} + \frac{y - x}{200}. \end{aligned}$$

Since x and y are distinct integers and $x < y$, we have $y - x \geq 1$, and we get

$$\Pr(W) \geq \frac{1}{2} + \frac{1}{200} = 0.505.$$

5.7 The Monty Hall Problem

The Monty Hall Problem is a well-known puzzle in probability theory. It is named after the host, Monty Hall, of the American television game show *Let's Make a Deal*. The problem became famous in 1990, when (part of) a reader's letter was published in Marilyn vos Savant's column *Ask Marilyn* in the magazine *Parade*:

> Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

Note that the host can always open a door that has a goat behind it. After the host has opened No. 3, we know that the car is either behind No. 1 or No. 2, and it seems that both these doors have the same probability (i.e.,

50%) of having the car behind them. We will prove below, however, that this is not true: It is indeed to our advantage to switch our choice.

We assume that the car is equally likely to be behind any of the three doors. Moreover, the host knows what is behind each door.

- We initially choose one of the three doors uniformly at random; this door remains closed.
- The host opens one of the other two doors that has a goat behind it.
- Our final choice is to switch to the other door that is still closed.

Let A be the event that we win the car and let B be the event that the initial door has a goat behind it. Then it is not difficult to see that event A occurs if and only if event B occurs. Therefore, the probability that we win the car is equal to

$$\Pr(A) = \Pr(B) = 2/3.$$

5.8 Conditional Probability

Anil Maheshwari has two children. We are told that one of them is a boy. What is the probability that the other child is also a boy? Most people will say that this probability is 1/2. We will show below that this is not the correct answer.

Since Anil has two children, the sample space is

$$S = \{(b, b), (b, g), (g, b), (g, g)\},$$

where, for example, (b, g) indicates that the youngest child is a boy and the oldest child is a girl. We assume a uniform probability function, so that each outcome has a probability of 1/4.

We are given the additional information that one of the two children is a boy, or, to be more precise, that at least one of the two children is a boy. This means that the actual sample space is not S, but

$$\{(b, b), (b, g), (g, b)\}.$$

When asking for the probability that the other child is also a boy, we are really asking for the probability that both children are boys. Since there is

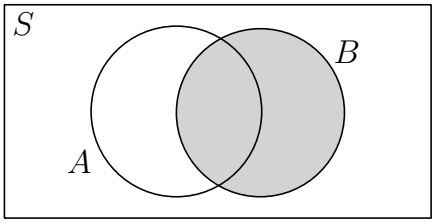
only one possibility (out of three) for both children to be boys, it follows that this probability is equal to 1/3.

This is an example of a *conditional probability*: We are asking for the probability of an event (both children are boys) given that another event (at least one of the two children is a boy) occurs.

Definition 5.8.1 Let $(S, \Pr)$ be a probability space and let A and B be two events with $\Pr(B) > 0$. The *conditional probability* $\Pr(A \mid B)$, pronounced as "the probability of A given B", is defined to be

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Let us try to understand where this definition comes from. Initially, the sample space is equal to S. When we are given the additional information that event B occurs, the sample space "shrinks" to B, and event A occurs if and only if event $A \cap B$ occurs.



You may think that $\Pr(A \mid B)$ should therefore be defined to be $\Pr(A \cap B)$. However, since the sum of all probabilities must be equal to 1, we have to normalize, i.e., divide by $\Pr(B)$. Equivalently, if $A = B$, we get $\Pr(A \mid A)$, which is the probability that event A occurs, given that event A occurs. This probability should be equal to 1. Indeed, using the definition, we do get

$$\Pr(A \mid A) = \frac{\Pr(A \cap A)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1.$$

In Exercise 5.10, you are asked to give a formal proof that our definition gives a valid probability function on the sample space S.

5.8.1 Anil's Children

Returning to Anil's two children, we saw that the sample space is

$$S = \{(b, b), (b, g), (g, b), (g, g)\}$$

and we assumed a uniform probability function. The events we considered are

$$A = \text{“both children are boys”}$$

and

$$B = \text{“at least one of the two children is a boy”},$$

and we wanted to know $\Pr(A \mid B)$. Writing A and B as subsets of the sample space S, we get

$$A = \{(b, b)\}$$

$$B = \{(b, b), (b, g), (g, b)\}.$$

It follows that

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)} = \frac{1/4}{3/4} = 1/3,$$

which is the same answer as we got before.

5.8.2 Rolling a Die

Assume we roll a fair die and consider the events

$$A = \text{“the result is 3”}$$

and

$$B = \text{“the result is odd”}.$$

Then

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)} = \frac{1/6}{3/6} = 1/3$$

and

$$\Pr(B \mid A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1.$$

This shows that, in general, $\Pr(A \mid B)$ is not equal to $\Pr(B \mid A)$.

Consider the event

$$C = \text{“the result is a prime number”},$$

i.e., the result is 2, 3, or 5. We have

$$\Pr(C \mid B) = \frac{\Pr(C \cap B)}{\Pr(B)} = \frac{2/6}{3/6} = 2/3$$

and

$$\Pr(C \mid A) = \frac{\Pr(C \cap A)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1.$$

Recall that $\overline{A}$ denotes the complement of A. Thus, this is the event

$$\overline{A} = \text{“the result is not 3”}.$$

We have

$$\Pr\left(C \mid \overline{A}\right) = \frac{\Pr\left(C \cap \overline{A}\right)}{\Pr\left(\overline{A}\right)} = \frac{2/6}{5/6} = 2/5.$$

This shows that, in general, $\Pr(C \mid A) + \Pr\left(C \mid \overline{A}\right)$ is not equal to 1. It should be an easy exercise to verify that, for the above events A and C, $\Pr(A \mid C) + \Pr\left(\overline{A} \mid C\right)$ is equal to 1. Intuitively, this should be true for any two events A and C with $\Pr(C) > 0$: When we are given that event C occurs, then either A occurs or A does not occur (in which case $\overline{A}$ occurs). The following lemma states that this intuition is indeed correct.

Lemma 5.8.2 *Let A and B be events with $\Pr(B) > 0$. Then*

$$\Pr(A \mid B) + \Pr\left(\overline{A} \mid B\right) = 1.$$

Proof. By definition, we have

$$\begin{aligned} \Pr(A \mid B) + \Pr\left(\overline{A} \mid B\right) &= \frac{\Pr(A \cap B)}{\Pr(B)} + \frac{\Pr\left(\overline{A} \cap B\right)}{\Pr(B)} \\ &= \frac{\Pr(A \cap B) + \Pr\left(\overline{A} \cap B\right)}{\Pr(B)}. \end{aligned}$$

Since the events $A \cap B$ and $\overline{A} \cap B$ are disjoint, we have, by Lemma 5.3.2,

$$\Pr(A \cap B) + \Pr\left(\overline{A} \cap B\right) = \Pr\left((A \cap B) \cup \left(\overline{A} \cap B\right)\right).$$

By drawing a Venn diagram, you will see that

$$(A \cap B) \cup \left(\overline{A} \cap B\right) = B,$$

implying that

$$\Pr(A \cap B) + \Pr\left(\overline{A} \cap B\right) = \Pr(B).$$

We conclude that

$$\Pr(A \mid B) + \Pr\left(\overline{A} \mid B\right) = \frac{\Pr(B)}{\Pr(B)} = 1.$$

■

5.9 The Law of Total Probability

Both Mick and Keith have a random birthday. What is the probability that they have the same birthday? We have seen in Section 5.5 that this probability is equal to 1/365. A common way to determine this probability is as follows: Consider Mick's birthday, which can be any of the 365 days of the year. By symmetry, it does not really matter what Mick's birthday is, so we just assume that it is July 26. Then Mick and Keith have the same birthday if and only Keith's birthday is also on July 26. Therefore, since Keith has a random birthday, the probability that Mick and Keith have the same birthday is equal to 1/365. The following theorem explains this reasoning.

Theorem 5.9.1 (Law of Total Probability) *Let $(S, \Pr)$ be a probability space and let A be an event. Assume that $B_1, B_2, \ldots, B_n$ is a sequence of events such that*

1. $\Pr(B_i) > 0$ *for all* i *with* $1 \leq i \leq n$,
2. *the events $B_1, B_2, \ldots, B_n$ are pairwise disjoint, and*
3. $\bigcup_{i=1}^{n} B_i = S$.

Then

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i) \cdot \Pr(B_i).$$

Proof. The assumptions imply that

$$\begin{aligned} A &= A \cap S \\ &= A \cap \left(\bigcup_{i=1}^{n} B_i \right) \\ &= \bigcup_{i=1}^{n} (A \cap B_i) \,. \end{aligned}$$

Since the events $A \cap B_1, A \cap B_2, \ldots, A \cap B_n$ are pairwise disjoint, it follows from Lemma 5.3.2 that

$$\begin{aligned} \Pr(A) &= \Pr\left(\bigcup_{i=1}^{n} (A \cap B_i) \right) \\ &= \sum_{i=1}^{n} \Pr\left(A \cap B_i\right) . \end{aligned}$$

The theorem follows by observing that, from Definition 5.8.1,

$$\Pr\left(A \cap B_i\right) = \Pr\left(A \mid B_i\right) \cdot \Pr\left(B_i\right) .$$

∎

Let us consider the three conditions in this theorem. The first condition is that $\Pr\left(B_i\right) > 0$, i.e., there is a positive probability that event B_i occurs. The second and third conditions, i.e.,

- the events $B_1, B_2, \ldots, B_n$ are pairwise disjoint, and
- $\bigcup_{i=1}^{n} B_i = S$,

are equivalent to

- exactly one of the events $B_1, B_2, \ldots, B_n$ is guaranteed to occur.

In the example in the beginning of this section, we wanted to know $\Pr(A)$, where A is the event

$$A = \text{“Mick and Keith have the same birthday”}.$$

In order to apply Theorem 5.9.1, we *define* a sequence $B_1, B_2, \ldots$ of events that satisfy the conditions in this theorem and for which $\Pr\left(A \mid B_i\right)$ is easy

to determine. For this example, we define the event B_i, for each i with $1 \leq i \leq 365$, to be

$$B_i = \text{“Mick's birthday is on the } i\text{-th day of the year”}.$$

It is clear that $\Pr(B_i) = 1/365 > 0$ and exactly one of the events $B_1, B_2, \ldots, B_{365}$ is guaranteed to occur. It follows that

$$\Pr(A) = \sum_{i=1}^{365} \Pr(A \mid B_i) \cdot \Pr(B_i).$$

To determine $\Pr(A \mid B_i)$, we assume that the event B_i occurs, i.e., we fix Mick's birthday to be the i-th day of the year. Given this event B_i, event A occurs if and only if Keith's birthday is also on the i-th day. Thus, we have $\Pr(A \mid B_i) = 1/365$ and it follows that

$$\begin{aligned} \Pr(A) &= \sum_{i=1}^{365} (1/365) \cdot \Pr(B_i) \\ &= (1/365) \sum_{i=1}^{365} \Pr(B_i) \\ &= (1/365) \cdot 1 \\ &= 1/365, \end{aligned}$$

which is the same answer as we got in the beginning of this section.

5.9.1 Flipping a Coin and Rolling Dice

Consider the following experiment:

- We flip a fair coin.
  - If the coin comes up heads, then we roll a fair die. Let R denote the result of this die.
  - If the coin comes up tails, then we roll two fair dice. Let R denote the sum of the results of these dice.

What is the probability that the value of R is equal to 2? That is, if we define the event A to be

$$A = \text{“the value of } R \text{ is equal to 2”},$$

then we want to know $\Pr(A)$. Since the value of R depends on whether the coin comes up heads or tails, we define the event

$$B = \text{“the coin comes up heads”}.$$

Since (i) both B and its complement $\overline{B}$ occur with a positive probability and (ii) exactly one of B and $\overline{B}$ is guaranteed to occur, we can apply Theorem 5.9.1 and get

$$\Pr(A) = \Pr(A \mid B) \cdot \Pr(B) + \Pr\left(A \mid \overline{B}\right) \cdot \Pr\left(\overline{B}\right).$$

We determine the four terms on the right-hand side:

- It should be clear that

$$\Pr(B) = \Pr\left(\overline{B}\right) = 1/2.$$

- To determine $\Pr(A \mid B)$, we assume that the event B occurs, i.e., the coin comes up heads. Because of this assumption, we roll one die, and the event A occurs if and only if the result of this roll is 2. It follows that

$$\Pr(A \mid B) = 1/6.$$

- To determine $\Pr\left(A \mid \overline{B}\right)$, we assume that the event $\overline{B}$ occurs, i.e., the coin comes up tails. Because of this assumption, we roll two dice, and the event A occurs if and only if both rolls result in 1. Since there are 36 possible outcomes when rolling two dice, it follows that

$$\Pr\left(A \mid \overline{B}\right) = 1/36.$$

We conclude that

$$\begin{aligned} \Pr(A) &= \Pr(A \mid B) \cdot \Pr(B) + \Pr\left(A \mid \overline{B}\right) \cdot \Pr\left(\overline{B}\right) \\ &= 1/6 \cdot 1/2 + 1/36 \cdot 1/2 \\ &= 7/72. \end{aligned}$$

5.10 Independent Events

Consider two events A and B with $\Pr(A) > 0$ and $\Pr(B) > 0$. We would like to define the notion of these two events being "independent". Intuitively, this should express that (i) whether or not event A occurs does not depend on whether or not event B occurs and (ii) whether or not event B occurs does not depend on whether or not event A occurs. In other words, (i) $\Pr(A)$ should be equal to the conditional probability $\Pr(A \mid B)$ and (ii) $\Pr(B)$ should be equal to conditional probability $\Pr(B \mid A)$. As we will show below, the following definition exactly captures this.

Definition 5.10.1 Let $(S, \Pr)$ be a probability space and let A and B be two events. We say that A and B are *independent* if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

Note that in this definition, it is not assumed that $\Pr(A) > 0$ and $\Pr(B) > 0$. If $\Pr(B) > 0$, then

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)},$$

and A and B are independent if and only if

$$\Pr(A \mid B) = \Pr(A).$$

Similarly, if $\Pr(A) > 0$, then A and B are independent if and only if

$$\Pr(B \mid A) = \Pr(B).$$

5.10.1 Rolling Two Dice

Assume we roll two dice; thus, the sample space is

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\},$$

where i is the result of the first die and j is the result of the second die. We assume a uniform probability function. Thus, each outcome has a probability of $1/36$.

Let D_1 denote the result of the first die and let D_2 denote the result of the second die. Consider the events

$$A = \text{“}D_1 + D_2 = 7\text{”}$$

and

$$B = \text{“}D_1 = 4\text{”}.$$

Are these events independent?

- Since
$$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$
we have $\Pr(A) = 6/36 = 1/6$.
- Since
$$B = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\},$$
we have $\Pr(B) = 6/36 = 1/6$.
- Since
$$A \cap B = \{(4,3)\},$$
we have $\Pr(A \cap B) = 1/36$.
- It follows that $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ and we conclude that A and B are independent.

As an exercise, you should verify that the events

$$A' = \text{“}D_1 + D_2 = 11\text{”}$$

and

$$B' = \text{“}D_1 = 5\text{”}$$

are not independent.

5.10.2 A Basic Property of Independent Events

Consider two events A and B. If these events are independent, then whether or not A occurs does not depend on whether or not B occurs. Since whether or not B occurs is the same as whether the complement $\overline{B}$ does not or does occur, it should not be a surprise that the events A and $\overline{B}$ are independent as well. The following lemma states that this is indeed the case.

Lemma 5.10.2 *If A and B are independent events, then A and $\overline{B}$ are also independent.*

Proof. To prove that A and $\overline{B}$ are independent, we have to show that

$$\Pr\left(A \cap \overline{B}\right) = \Pr(A) \cdot \Pr\left(\overline{B}\right).$$

Using Lemma 5.3.3, this is equivalent to showing that

$$\Pr\left(A \cap \overline{B}\right) = \Pr(A) \cdot (1 - \Pr(B)). \tag{5.3}$$

Since the events $A \cap B$ and $A \cap \overline{B}$ are disjoint and

$$A = (A \cap B) \cup \left(A \cap \overline{B}\right),$$

it follows from Lemma 5.3.2 that

$$\Pr(A) = \Pr(A \cap B) + \Pr\left(A \cap \overline{B}\right).$$

Since A and B are independent, we have

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

It follows that

$$\Pr(A) = \Pr(A) \cdot \Pr(B) + \Pr\left(A \cap \overline{B}\right),$$

which is equivalent to (5.3). ∎

5.10.3 Pairwise and Mutually Independent Events

We have defined the notion of two events being independent. The following definition generalizes this in two ways to sequences of events:

Definition 5.10.3 Let $(S, \Pr)$ be a probability space, let $n \geq 2$, and let $A_1, A_2, \ldots, A_n$ be a sequence of events.

1. We say that this sequence is *pairwise independent* if for any two distinct indices i and j, the events A_i and A_j are independent, i.e.,

$$\Pr(A_i \cap A_j) = \Pr(A_i) \cdot \Pr(A_j).$$

2. We say that this sequence is *mutually independent* if for all k with $2 \leq k \leq n$ and all $i_1 < i_2 < \ldots < i_k$,

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

Thus, in order to show that the sequence $A_1, A_2, \ldots, A_n$ is pairwise independent, we have to verify $\binom{n}{2}$ equalities. On the other hand, to show that this sequence is mutually independent, we have to verify $\sum_{k=2}^{n} \binom{n}{k} = 2^n - 1 - n$ equalities.

For example, if we want to prove that the sequence A, B, C of three events is mutually independent, then we have to show that

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B),$$
$$\Pr(A \cap C) = \Pr(A) \cdot \Pr(C),$$
$$\Pr(B \cap C) = \Pr(B) \cdot \Pr(C),$$

and

$$\Pr(A \cap B \cap C) = \Pr(A) \cdot \Pr(B) \cdot \Pr(C).$$

To give an example, consider flipping three coins and assume that the result is a uniformly random element from the sample space

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\},$$

where, e.g., HHT indicates that the first two coins come up heads and the third coin comes up tails. For $i = 1, 2, 3$, let f_i denote the result of the i-th flip, and define the events

$$A = \text{“}f_1 = f_2\text{”},$$
$$B = \text{“}f_2 = f_3\text{”},$$

and

$$C = \text{“}f_1 = f_3\text{”}.$$

If we write these events as subsets of the sample space, then we get

$$A = \{HHH, HHT, TTH, TTT\},$$
$$B = \{HHH, THH, HTT, TTT\},$$

and

$$C = \{HHH, HTH, THT, TTT\}.$$

It follows that

$$\begin{array}{rcrcrcl}
\Pr(A) & = & |A|/|S| & = & 4/8 & = & 1/2, \\
\Pr(B) & = & |B|/|S| & = & 4/8 & = & 1/2, \\
\Pr(C) & = & |C|/|S| & = & 4/8 & = & 1/2, \\
\Pr(A \cap B) & = & |A \cap B|/|S| & = & 2/8 & = & 1/4, \\
\Pr(A \cap C) & = & |A \cap C|/|S| & = & 2/8 & = & 1/4, \\
\Pr(B \cap C) & = & |B \cap C|/|S| & = & 2/8 & = & 1/4.
\end{array}$$

Thus, the sequence A, B, C is pairwise independent. Since

$$\Pr(A \cap B \cap C) = |A \cap B \cap C|/|S| = 2/8 = 1/4,$$

we have

$$\Pr(A \cap B \cap C) \neq \Pr(A) \cdot \Pr(B) \cdot \Pr(C)$$

and, therefore, the sequence A, B, C is not mutually independent. Of course, this is not surprising: If both events A and B occur, then event C also occurs.

5.11 Describing Events by Logical Propositions

We have defined an event to be a subset of a sample space. In several examples, however, we have described events in plain English or as logical propositions.

- Since the intersection ($\cap$) of sets corresponds to the conjunction ($\wedge$) of propositions, we often write $A \wedge B$ for the event "both A and B occur".
- Similarly, since the union ($\cup$) of sets corresponds to the disjunction ($\vee$) of propositions, we often write $A \vee B$ for the event "A or B occurs".

5.11.1 Flipping a Coin and Rolling a Die

If we flip a coin and roll a die, the sample space is

$$S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}.$$

The events

$$A = \text{"the coin comes up heads"}$$

and

$$B = \text{"the result of the die is even"}$$

correspond to the subsets

$$A = \{H1, H2, H3, H4, H5, H6\}$$

and

$$B = \{H2, H4, H6, T2, T4, T6\}$$

of the sample space S, respectively. The event that both A and B occur is written as $A \wedge B$ and corresponds to the subset

$$A \cap B = \{H2, H4, H6\}$$

of S. The event that A or B occurs is written as $A \vee B$ and corresponds to the subset

$$A \cup B = \{H1, H2, H3, H4, H5, H6, T2, T4, T6\}$$

of S.

Assume that both the coin and the die are fair and the result of rolling the die is independent of the result of flipping the coin. The probability that both A and B occur, i.e., $\Pr(A \wedge B)$, is equal to $|A \cap B|/|S| = 3/12 = 1/4$. We can also use independence to determine this probability:

$$\Pr(A \wedge B) = \Pr(A) \cdot \Pr(B) = (1/2)(3/6) = 1/4.$$

Observe that when we determine $\Pr(A)$, we do not consider the entire sample space S. Instead, we consider the coin's sample space, which is $\{H, T\}$. Similarly, when we determine $\Pr(B)$, we consider the die's sample space, which is $\{1, 2, 3, 4, 5, 6\}$.

The probability that A or B occurs, i.e., $\Pr(A \vee B)$, is equal to $|A \cup B|/|S| = 9/12 = 3/4$.

5.11.2 Flipping Coins

Let $n \geq 1$ be an integer and assume we flip n fair coins. For each i with $1 \leq i \leq n$, define the event

$$A_i = \text{“the } i\text{-th coin comes up heads”}.$$

We assume that the coin flips are independent of each other, by which we mean that the sequence $A_1, A_2, \ldots, A_n$ of events is mutually independent. Define the event

$$A = A_1 \wedge A_2 \wedge \cdots \wedge A_n.$$

What is $\Pr(A)$, i.e., the probability that all n coins come up heads? Since there are 2^n many possible outcomes for n coin flips and only one of them

satisfies event A, this probability is equal to $1/2^n$. Alternatively, we can use independence to determine $\Pr(A)$:

$$\begin{aligned}\Pr(A) &= \Pr(A_1 \wedge A_2 \wedge \cdots \wedge A_n) \\ &= \Pr(A_1) \cdot \Pr(A_2) \cdots \Pr(A_n).\end{aligned}$$

Since each coin is fair, we have $\Pr(A_i) = 1/2$ and, thus, we get

$$\Pr(A) = \underbrace{(1/2)(1/2)\cdots(1/2)}_{n \text{ times}} = 1/2^n.$$

5.11.3 The Probability of a Circuit Failing

Consider a circuit C that consists of n components $C_1, C_2, \ldots, C_n$. Let p be a real number with $0 < p < 1$ and assume that any component fails with probability p, independently of the other components. For each i with $1 \leq i \leq n$, define the event

$$A_i = \text{“component } C_i \text{ does not fail”}.$$

Let A be the event

$$A = \text{“the entire circuit does not fail”}.$$

- Assume that the entire circuit fails when at least one component fails. What is $\Pr(A)$, i.e., the probability that the circuit does not fail? Observe that, by our assumption,

  $$A = A_1 \wedge A_2 \wedge \cdots \wedge A_n.$$

  Using independence and Lemma 5.10.2, we get

  $$\begin{aligned}\Pr(A) &= \Pr(A_1 \wedge A_2 \wedge \cdots \wedge A_n) \\ &= \Pr(A_1) \cdot \Pr(A_2) \cdots \Pr(A_n) \\ &= \underbrace{(1-p)(1-p)\cdots(1-p)}_{n \text{ times}} \\ &= (1-p)^n.\end{aligned}$$

  Since $0 < p < 1$, we have $\lim_{n\to\infty} \Pr(A) = 0$. We conclude that for large values of n, it is very likely that the circuit fails.

- Now assume that the entire circuit fails when all components fail. Again, we want to know the probability $\Pr(A)$ that the circuit does not fail. In this case, we have

$$\overline{A} = \overline{A}_1 \wedge \overline{A}_2 \wedge \cdots \wedge \overline{A}_n,$$

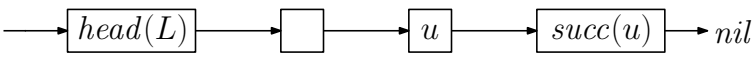
and we get

$$\begin{aligned}
\Pr(A) &= 1 - \Pr\left(\overline{A}\right) \\
&= 1 - \Pr\left(\overline{A}_1 \wedge \overline{A}_2 \wedge \cdots \wedge \overline{A}_n\right) \\
&= 1 - \Pr\left(\overline{A}_1\right) \cdot \Pr\left(\overline{A}_2\right) \cdots \Pr\left(\overline{A}_n\right) \\
&= 1 - \underbrace{p \cdot p \cdots p}_{n \text{ times}} \\
&= 1 - p^n.
\end{aligned}$$

Since $0 < p < 1$, we have $\lim_{n\to\infty} \Pr(A) = 1$. Thus, for large values of n, it is very likely that the circuit does not fail.

5.12 Choosing a Random Element in a Linked List

Consider a linked list L. Each node u in L stores a pointer to its successor node $succ(u)$. If u is the last node in L, then u does not have a successor and $succ(u) = nil$. We are also given a pointer to the first node $head(L)$ of L.



Our task is to choose, uniformly at random, a node in L. Thus, if this list has n nodes, then each node must have a probability of $1/n$ of being chosen.

We assume that we are given a function RANDOM: For any integer $i \geq 1$, a call to RANDOM(i) returns a uniformly random element from the set $\{1, 2, \ldots, i\}$; the value returned is independent of previous calls to this function.

To make the problem interesting, we assume that we do not know the value of n, i.e., at the start, we do not know the number of nodes in the list L. Also, we are allowed to only make one pass over this list. We will prove below that the following algorithm solves the problem:

Algorithm ChooseRandomNode(L)**:**

```
u = head(L);
i = 1;
while u ≠ nil
do r = Random(i);
    if r = 1
    then x = u
    endif;
    u = succ(u);
    i = i + 1
endwhile;
return x
```

In one iteration of the while-loop, the call to Random(i) returns a uniformly random element r from the set $\{1, 2, \ldots, i\}$. If $r = 1$, which happens with probability $1/i$, the value of x is set to the currently visited node. Thus,

- in the first iteration, x is set to the first node of L with probability 1,
- in the second iteration, x is set to the second node of L with probability $1/2$,
- in the third iteration, x is set to the third node of L with probability $1/3$,
- in the last iteration, x is set to the last node of L with probability $1/|L|$.

We now prove that the output x of algorithm ChooseRandomNode(L) is a uniformly random node of the list L. Let n denote the number of nodes in L and let v be an arbitrary node in L. Thus, we have to prove that, after the algorithm has terminated, $x = v$ with probability $1/n$.

Let k be the integer such that v is the k-th node in L; thus, $1 \leq k \leq n$. We observe that, after the algorithm has terminated, $x = v$ if and only if

- during the k-th iteration, the value of x is set to v, and
- for all $i = k + 1, k + 2, \ldots, n$, during the i-th iteration, the value of x does not change.

Define the event

$$A = \text{“after the algorithm has terminated, } x = v\text{”}.$$

For each i with $1 \leq i \leq n$, define the event

$$A_i = \text{“the value of } x \text{ changes during the } i\text{-th iteration”}.$$

Then

$$A = A_k \wedge \overline{A}_{k+1} \wedge \overline{A}_{k+2} \wedge \overline{A}_{k+3} \wedge \cdots \wedge \overline{A}_n.$$

Since the events $A_1, A_2, \ldots, A_n$ are mutually independent, it follows that

$$\begin{aligned}
\Pr(A) &= \Pr\left(A_k \wedge \overline{A}_{k+1} \wedge \overline{A}_{k+2} \wedge \overline{A}_{k+3} \wedge \cdots \wedge \overline{A}_n\right) \\
&= \Pr\left(A_k\right) \cdot \Pr\left(\overline{A}_{k+1}\right) \cdot \Pr\left(\overline{A}_{k+2}\right) \cdot \Pr\left(\overline{A}_{k+3}\right) \cdots \Pr\left(\overline{A}_n\right) \\
&= \frac{1}{k} \cdot \left(1 - \frac{1}{k+1}\right) \cdot \left(1 - \frac{1}{k+2}\right) \cdot \left(1 - \frac{1}{k+3}\right) \cdots \left(1 - \frac{1}{n}\right) \\
&= \frac{1}{k} \cdot \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdot \frac{k+2}{k+3} \cdots \frac{n-1}{n} \\
&= \frac{1}{n}.
\end{aligned}$$

5.13 Long Runs in Random Bitstrings

Let n be a large integer and assume we flip a fair coin n times, where all flips are mutually independent. If we write 0 for heads and 1 for tails, then we obtain a random bitstring

$$S = s_1 s_2 \ldots s_n.$$

A *run of length* k is a consecutive subsequence of S, all of whose bits are the same. For example, the bitstring

$$00111100101000011000$$

contains, among others, the following consecutive subsequences in bold,

$$00\mathbf{1111}00101\mathbf{00}0\mathbf{1}1000,$$

which are runs of lengths 4, 2, and 1, respectively.

Would you be surprised to see a "long" run in the random bitstring S, say a run of length about $\log n$? Most people will answer this question with "yes". We will prove below, however, that the correct answer is "no": The probability that this happens is (about) $1 - 1/n$; thus, it converges to 1 when n goes to infinity. In other words, you should be surprised if a random bitstring does *not* contain a run of length about $\log n$.

We choose a positive integer k and define the event

$$A = \text{“}S \text{ contains a run of length at least } k\text{”}.$$

We are going to prove a lower bound on $\Pr(A)$ in terms of n and k. At the end, we will show that by taking k to be slightly less than $\log n$, we have $\Pr(A) \geq 1 - 1/n$.

For each i with $1 \leq i \leq n - k + 1$, we define the event

$$A_i = \text{“the subsequence of length } k \text{ starting at position } i \text{ is a run”}.$$

Since

$$A = A_1 \vee A_2 \vee \cdots \vee A_{n-k+1},$$

we have

$$\Pr(A) = \Pr\left(A_1 \vee A_2 \vee \cdots \vee A_{n-k+1}\right).$$

Since the probability on the right-hand side is difficult to analyze, we consider the complement of the event A, i.e., the event

$$\overline{A} = \text{“each run in } S \text{ has length less than } k\text{”}.$$

Note that

$$\overline{A} = \overline{A}_1 \wedge \overline{A}_2 \wedge \cdots \wedge \overline{A}_{n-k+1},$$

where $\overline{A}_i$ is the complement of A_i, i.e, the event

$$\overline{A}_i = \text{“the subsequence of length } k \text{ starting at position } i \text{ is not a run.”}$$

It follows that

$$\Pr\left(\overline{A}\right) = \Pr\left(\overline{A}_1 \wedge \overline{A}_2 \wedge \cdots \wedge \overline{A}_{n-k+1}\right). \tag{5.4}$$

We determine $\Pr\left(\overline{A}_i\right)$, by first determining $\Pr\left(A_i\right)$. The event A_i occurs if and only if

$$s_i = s_{i+1} = \cdots = s_{i+k-1} = 0$$

or

$$s_i = s_{i+1} = \cdots = s_{i+k-1} = 1.$$

Since the coin flips are mutually independent, it follows that

$$\Pr(A_i) = 1/2^k + 1/2^k = 1/2^{k-1}$$

and, therefore,

$$\Pr\left(\overline{A}_i\right) = 1 - \Pr(A_i) = 1 - 1/2^{k-1}.$$

Is the probability on the right-hand side of (5.4) equal to the product of the individual probabilities? If the events $\overline{A}_1, \overline{A}_2, \ldots, \overline{A}_{n-k+1}$, are mutually independent, then the answer is "yes". However, it should be clear that, for example, the events $\overline{A}_1$ and $\overline{A}_2$ are not independent: If we are told that event A_1 occurs, then the first k bits in the sequence S are equal; let us say they are all equal to 0. In this case, the probability that event A_2 occurs is equal to the probability that the $(k+1)$-st bit in S is 0, which is equal to $1/2$ and not $1/2^{k-1}$. So it seems that we are stuck. However, there is a way out:

Let us assume that the integer k is chosen such that n/k is an integer. We divide the sequence $S = s_1 s_2 \ldots s_n$ into n/k *blocks*, each having length k. Thus,

- the first block is the subsequence $s_1 s_2 \ldots s_k$,
- the second block is the subsequence $s_{k+1} s_{k+2} \ldots s_{2k}$,
- the third block is the subsequence $s_{2k+1} s_{2k+2} \ldots s_{3k}$,
- the (n/k)-th block is the subsequence $s_{n-k+1} s_{n-k+2} \ldots s_n$.

For each i with $1 \leq i \leq n/k$, we define the event

$$B_i = \text{"the } i\text{-th block is a run"}.$$

Thus, the complement of B_i is the event

$$\overline{B}_i = \text{"the } i\text{-th block is not a run"}.$$

Since $B_i = A_{(i-1)k+1}$ and $\overline{B}_i = \overline{A}_{(i-1)k+1}$, we have

$$\Pr\left(\overline{B}_i\right) = 1 - 1/2^{k-1}.$$

Observe that

- the events $\overline{B}_1, \overline{B}_2, \ldots, \overline{B}_{n/k}$ are mutually independent, because the blocks do not overlap, and
- if the event $\overline{A}$ occurs, then the event $\overline{B}_1 \wedge \overline{B}_2 \wedge \cdots \wedge \overline{B}_{n/k}$ also occurs.

Using Lemma 5.3.6, it follows that

$$\begin{aligned}
\Pr\left(\overline{A}\right) &\leq \Pr\left(\overline{B}_1 \wedge \overline{B}_2 \wedge \cdots \wedge \overline{B}_{n/k}\right) \\
&= \Pr\left(\overline{B}_1\right) \cdot \Pr\left(\overline{B}_2\right) \cdots \Pr\left(\overline{B}_{n/k}\right) \\
&= \left(1 - 1/2^{k-1}\right) \cdot \left(1 - 1/2^{k-1}\right) \cdots \left(1 - 1/2^{k-1}\right) \\
&= \left(1 - 1/2^{k-1}\right)^{n/k}.
\end{aligned}$$

Using the inequality $1 - x \leq e^{-x}$, see (5.2), we get

$$1 - 1/2^{k-1} \leq e^{-1/2^{k-1}} = e^{-2/2^k}$$

and, thus,

$$\Pr\left(\overline{A}\right) \leq \left(e^{-2/2^k}\right)^{n/k} = e^{-2n/(k2^k)}. \tag{5.5}$$

Note that until now, k was arbitrary. We choose k to be

$$k = \log n - 2 \log\log n.$$

Using basic properties of logarithms, see Section 2.4, we will show below that, for this choice of k, the right-hand side in (5.5) is a "nice" function of n.

In Section 2.4, we have seen that

$$2^{\log n} = n$$

and

$$2^{2 \log\log n} = \log^2 n.$$

It follows that

$$2^k = 2^{\log n - 2\log\log n} = \frac{2^{\log n}}{2^{2\log\log n}} = \frac{n}{\log^2 n}.$$

Thus,

$$\begin{aligned}
\frac{2n}{k2^k} &= \frac{2\log^2 n}{k} \\
&= \frac{2\log^2 n}{\log n - 2\log\log n} \\
&\geq \frac{2\log^2 n}{\log n} \\
&= 2\log n \\
&= 2\frac{\ln n}{\ln 2} \\
&\geq \ln n,
\end{aligned}$$

implying that

$$\begin{aligned}
\Pr\left(\overline{A}\right) &\leq e^{-2n/(k2^k)} \\
&\leq e^{-\ln n} \\
&= 1/n.
\end{aligned}$$

We conclude that, for the value of k chosen above,

$$\Pr(A) = 1 - \Pr\left(\overline{A}\right) \geq 1 - 1/n.$$

Thus, with probability at least $1 - 1/n$, a random bitstring of length n contains a run of length at least $\log n - 2\log\log n$.

We remark that we cheated a bit here, because we assumed that both k and n/k are integers. Assume that n is of the form 2^{2^m}, for some positive integer m. Then both $\log n$ and $\log\log n$ are integers and, thus, k is an integer as well. In a correct derivation, we divide the sequence S into $\lfloor n/k \rfloor$ blocks of size k and, if n/k is not an integer, one block of length less than k. We then get

$$\begin{aligned}
\Pr\left(\overline{A}\right) &\leq \left(1 - 1/2^{k-1}\right)^{\lfloor n/k \rfloor} \\
&\leq \left(e^{-2/2^k}\right)^{\lfloor n/k \rfloor} \\
&= e^{-2\lfloor n/k \rfloor/2^k}.
\end{aligned}$$

As we have seen before, for $k = \log n - 2\log\log n$, we have $2^k = n/\log^2 n$. Since

$$\lfloor n/k \rfloor > n/k - 1,$$

we get

$$\begin{aligned}\frac{2\lfloor n/k \rfloor}{2^k} &> \frac{2(n/k-1)}{2^k} \\ &= \frac{(2\log^2 n)(n/k-1)}{n} \\ &= \frac{2\log^2 n}{k} - \frac{2\log^2 n}{n} \\ &\geq \ln n - \frac{2\log^2 n}{n}\end{aligned}$$

and, thus,

$$\begin{aligned}\Pr\left(\overline{A}\right) &\leq e^{-2\lfloor n/k \rfloor/2^k} \\ &\leq e^{-\ln n + (2\log^2 n)/n} \\ &= e^{-\ln n} \cdot e^{(2\log^2 n)/n} \\ &= (1/n) \cdot \left(1 + O((\log^2 n)/n)\right) \\ &= 1/n + O\left((\log^2 n)/n^2\right).\end{aligned}$$

This upper bound is larger than the upper bound we had before by only a small additive factor of $O((\log^2 n)/n^2)$.

5.14 Infinite Probability Spaces

In Section 5.2, we defined a sample space to be any non-empty countable set. All sample spaces that we have seen so far are finite. In some cases, infinite (but countable) sample spaces arise in a natural way. To give an example, assume we flip a fair coin repeatedly and independently until it comes up heads for the first time. The sample space S is the set of sequences of all coin flips that can occur. If we denote by T^nH the sequence consisting of n tails followed by one heads, then

$$\begin{aligned}S &= \{H, TH, TTH, TTTH, TTTTH, \ldots\} \\ &= \{T^nH : n \geq 0\},\end{aligned}$$

which is indeed an infinite set.

Since the coin is fair and the coin flips are mutually independent, the outcome T^nH has a probability of $(1/2)^{n+1}$, i.e.,

$$\Pr\left(T^nH\right) = (1/2)^{n+1}.$$

Recall that according to Definition 5.2.2, in order for this to be a valid probability function, the sum of all probabilities must be equal to 1, i.e., the infinite series

$$\sum_{n=0}^{\infty} \Pr\left(T^nH\right) = \sum_{n=0}^{\infty} (1/2)^{n+1}$$

must be equal to 1. Since you may have forgotten about infinite series, we recall the definition in the following subsection.

5.14.1 Infinite Series

Definition 5.14.1 Let $a_0, a_1, a_2, \ldots$ be an infinite sequence of real numbers. If

$$\lim_{N\to\infty} \sum_{n=0}^{N} a_n = \lim_{N\to\infty} \left(a_0 + a_1 + a_2 + \cdots + a_N\right)$$

exists, then we say that the infinite series $\sum_{n=0}^{\infty} a_n$ *converges*. In this case, the value of this infinite series is equal to

$$\sum_{n=0}^{\infty} a_n = \lim_{N\to\infty} \sum_{n=0}^{N} a_n.$$

For example, let x be a real number with $x \neq 1$, and define $a_n = x^n$ for $n \geq 0$. We claim that

$$\sum_{n=0}^{N} a_n = \sum_{n=0}^{N} x^n = 1 + x + x^2 + \cdots + x^N = \frac{1 - x^{N+1}}{1 - x},$$

which can be proved either by induction on N or by verifying that

$$(1 - x)\left(1 + x + x^2 + \cdots + x^N\right) = 1 - x^{N+1}.$$

If $-1 < x < 1$, then $\lim_{N\to\infty} x^{N+1} = 0$ and it follows that

$$\begin{aligned}\sum_{n=0}^{\infty} x^n &= \lim_{N\to\infty} \sum_{n=0}^{N} x^n \\ &= \lim_{N\to\infty} \frac{1-x^{N+1}}{1-x} \\ &= \frac{1}{1-x}.\end{aligned}$$

We have proved the following result:

Lemma 5.14.2 *If x is a real number with $-1 < x < 1$, then*

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

Now we can return to the coin flipping example that we saw in the beginning of Section 5.14. If we take $x = 1/2$ in Lemma 5.14.2, then we get

$$\begin{aligned}\sum_{n=0}^{\infty} \Pr\left(T^n H\right) &= \sum_{n=0}^{\infty} (1/2)^{n+1} \\ &= (1/2) \sum_{n=0}^{\infty} (1/2)^n \\ &= (1/2) \cdot \frac{1}{1-1/2} \\ &= 1.\end{aligned}$$

Thus, Pr is indeed a valid probability function on the infinite sample space $S = \{T^n H : n \geq 0\}$.

We have seen in Lemma 5.14.2 that the infinite series $\sum_{n=0}^{\infty} x^n$ converges if $-1 < x < 1$. It is not difficult to see that for all other values of x, the limit

$$\lim_{N\to\infty} \sum_{n=0}^{N} x^n$$

does not exist. As a result, if $x \geq 1$ or $x \leq -1$, the infinite series $\sum_{n=0}^{\infty} x^n$ does not converge. Another example of an infinite series that does not converge is

$$\sum_{n=1}^{\infty} 1/n = 1 + 1/2 + 1/3 + 1/4 + \cdots$$

In Section 6.8.3, we will prove that

$$\sum_{n=1}^{N} 1/n = 1 + 1/2 + 1/3 + 1/4 + \cdots + 1/N$$

is about $\ln N$. It follows that

$$\lim_{N\to\infty} \sum_{n=0}^{N} 1/n$$

is about

$$\lim_{N\to\infty} \ln N,$$

which clearly does not exist.

5.14.2 Who Flips the First Heads

Consider a game in which two players P_1 and P_2 take turns flipping, independently, a fair coin. Thus, first P_1 flips the coin, then P_2 flips the coin, then P_1 flips the coin, then P_2 flips the coin, etc. The player who flips heads first is the winner of the game.

Who is more likely to win this game? Our intuition says that P_1 has an advantage, because he is the player who starts: If the first flip is heads, then the game is over and P_1 wins. We will prove below that this intuition is correct: P_1 has a probability of 2/3 of winning the game and, thus, the winning probability of P_2 is only 1/3.

The sample space S is the set of sequences of all coin flips that can occur. Since the game is over as soon as a heads is flipped, we have

$$S = \{T^n H : n \geq 0\}.$$

Since P_1 starts, the event

$$A = \text{“}P_1 \text{ wins the game”}$$

corresponds to the subset

$$A = \{T^n H : n \geq 0 \text{ and } n \text{ is even}\},$$

which we rewrite as

$$A = \{T^{2m} H : m \geq 0\}.$$

The probability that P_1 wins the game is equal to

$$\begin{aligned}\Pr(A) &= \sum_{m=0}^{\infty} \Pr\left(T^{2m}H\right) \\ &= \sum_{m=0}^{\infty} (1/2)^{2m+1} \\ &= (1/2)\sum_{m=0}^{\infty} (1/2)^{2m} \\ &= (1/2)\sum_{m=0}^{\infty} (1/4)^{m}.\end{aligned}$$

By taking $x = 1/4$ in Lemma 5.14.2, we get

$$\Pr(A) = (1/2) \cdot \frac{1}{1 - 1/4} = 2/3.$$

Let B be the event

$$B = \text{“}P_2 \text{ wins the game”}.$$

Since either P_1 or P_2 wins the game, we have

$$\Pr(B) = 1 - \Pr(A) = 1 - 2/3 = 1/3.$$

Let us verify, using an infinite series, that $\Pr(B)$ is indeed equal to $1/3$. The event B corresponds to the subset

$$B = \{T^n H : n \geq 0 \text{ and } n \text{ is odd}\},$$

which we rewrite as

$$B = \{T^{2m+1}H : m \geq 0\}.$$

The probability that P_2 wins the game is thus equal to

$$\begin{aligned}
\Pr(B) &= \sum_{m=0}^{\infty} \Pr\left(T^{2m+1}H\right) \\
&= \sum_{m=0}^{\infty} (1/2)^{2m+2} \\
&= (1/4)\sum_{m=0}^{\infty} (1/2)^{2m} \\
&= (1/4)\sum_{m=0}^{\infty} (1/4)^{m} \\
&= (1/4)\cdot\frac{1}{1-1/4} = 1/3.
\end{aligned}$$

5.14.3 Who Flips the Second Heads

Let us change the game from the previous subsection: Again, the two players P_1 and P_2 take turns flipping, independently, a fair coin, where P_1 starts. The game ends as soon as a second heads comes up. The player who flips the second heads wins the game.

In this game, a sequence of coin flips can occur if and only if (i) the sequence contains exactly two heads and (ii) the last element in the sequence is heads. Thus, the sample space S is given by

$$S = \{T^m H T^n H : m \geq 0, n \geq 0\}.$$

The event

$$A = \text{“}P_1 \text{ wins the game”}$$

corresponds to the subset

$$A = \{T^m H T^n H : m \geq 0, n \geq 0,\ m+n \text{ is odd}\}.$$

We split this event into two events

$$A_1 = \text{“}P_1 \text{ flips both the first and the second heads”}$$

and

$$A_2 = \text{“}P_2 \text{ flips the first heads and } P_1 \text{ flips the second heads”}.$$

If we write these two events as subsets of the sample space S, we get

$$\begin{aligned} A_1 &= \{T^m HT^n H : m \geq 0, n \geq 0,\ m \text{ is even and } n \text{ is odd}\} \\ &= \{T^{2k} HT^{2\ell+1} H : k \geq 0, \ell \geq 0\} \end{aligned}$$

and

$$\begin{aligned} A_2 &= \{T^m HT^n H : m \geq 0, n \geq 0,\ m \text{ is odd and } n \text{ is even}\} \\ &= \{T^{2k+1} HT^{2\ell} H : k \geq 0, \ell \geq 0\}. \end{aligned}$$

Observe that $A_1 \cap A_2 = \emptyset$ and $A = A_1 \cup A_2$. We have

$$\begin{aligned} \Pr(A_1) &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \Pr\left(T^{2k} HT^{2\ell+1} H\right) \\ &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} (1/2)^{2k+2\ell+3} \\ &= (1/2^3)\sum_{k=0}^{\infty}(1/2)^{2k}\sum_{\ell=0}^{\infty}(1/2)^{2\ell} \\ &= (1/8)\sum_{k=0}^{\infty}(1/4)^{k}\sum_{\ell=0}^{\infty}(1/4)^{\ell} \\ &= (1/8)\sum_{k=0}^{\infty}(1/4)^{k}\cdot\frac{1}{1-1/4} \\ &= (1/6)\sum_{k=0}^{\infty}(1/4)^{k} \\ &= (1/6)\cdot\frac{1}{1-1/4} \\ &= 2/9 \end{aligned}$$

and

$$\begin{aligned} \Pr(A_2) &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} \Pr\left(T^{2k+1} HT^{2\ell} H\right) \\ &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} (1/2)^{2k+2\ell+3} \\ &= 2/9. \end{aligned}$$

Thus, the probability that P_1 wins the game is equal to

$$\begin{aligned} \Pr(A) &= \Pr(A_1) + \Pr(A_2) \\ &= 2/9 + 2/9 \\ &= 4/9. \end{aligned}$$

The probability that P_2 wins the game is equal to

$$1 - \Pr(A) = 5/9.$$

Thus, P_2 has a slightly larger probability of winning the game.

5.15 Exercises

5.1 Consider a coin that has 0 on one side and 1 on the other side. We flip this coin once and roll a die twice, and are interested in the product of the three numbers.

- What is the sample space?
- How many possible events are there?
- If both the coin and the die are fair, how would you define the probability function Pr for this sample space?

5.2 Consider the sample space $S = \{a, b, c, d\}$ and a probability function $\Pr : S \to \mathbb{R}$ on S. Define the events $A = \{a\}$, $B = \{a, b\}$, $C = \{a, b, c\}$, and $D = \{b, d\}$. You are given that $\Pr(A) = 1/10$, $\Pr(B) = 1/2$, and $\Pr(C) = 7/10$. Determine $\Pr(D)$.

5.3 Let n be a positive integer. We flip a fair coin $2n$ times and consider the possible outcomes, which are strings of length $2n$ with each character being H (= heads) or T (= tails). Thus, we take the sample space S to be the set of all such strings. Since our coin is fair, each string of S should have the same probability. Thus, we define $\Pr(s) = 1/|S|$ for each string s in S. In other words, we have a uniform probability space.

You are asked to determine the probability that in the sequence of $2n$ flips, the coin comes up heads exactly n times:

- What is the event A that describes this?

- Determine $\Pr(A)$.

5.4 A cup contains two pennies (P), one nickel (N), and one dime (D). You choose one coin uniformly at random, and then you choose a second coin from the remaining coins, again uniformly at random.

- Let S be the sample space consisting of all ordered pairs of letters P, N, and D that represent the possible outcomes. Write out all elements of S.

- Determine the probability for each element in this sample space.

5.5 Let $k \geq 2$ be an integer and consider the sample space S consisting of all sequences of k characters, where each character is one of the digits $0, 1, 2, \ldots, 9$.

If we choose a sequence s uniformly at random from the sample space S, what is the probability that none of the digits in s is equal to 5?

5.6 The Fibonacci numbers are defined as follows: $f_0 = 0$, $f_1 = 1$, and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$.

Let n be a large integer. A *Fibonacci die* is a die that has f_n faces. Such a die is fair: If we roll it, each face is on top with the same probability $1/f_n$. There are three different types of Fibonacci dice:

- D_1: f_{n-2} of its faces show the number 1 and the other f_{n-1} faces show the number 4.

- D_2: Each face shows the number 3.

- D_3: f_{n-2} of its faces show the number 5 and the other f_{n-1} faces show the number 2.

Assume we roll each of D_1, D_2, and D_3 once, independently of each other. Let R_1, R_2, and R_3 be the numbers on the top face of D_1, D_2, and D_3, respectively. Determine

$$\Pr(R_1 > R_2)$$

and

$$\Pr(R_2 > R_3),$$

and show that

$$\Pr(R_3 > R_1) = \frac{f_{n-2} f_{n+1}}{f_n^2}.$$

5.7 A group of ten people sits down, uniformly at random, around a table. Lindsay and Simon are part of this group. Determine the probability that Lindsay and Simon sit next to each other.

5.8 In Section 5.4.1, we have seen the different cards that are part of a standard deck of cards.

- You choose 2 cards uniformly at random from the 13 spades in a deck of 52 cards. Determine the probability that you choose an ace and a king.
- You choose 2 cards uniformly at random from a deck of 52 cards. Determine the probability that you choose an ace and a king.
- You choose 2 cards uniformly at random from a deck of 52 cards. Determine the probability that you choose an ace and a king of the same suit.

5.9 Prove the inequality in (5.2), i.e., prove that

$$1 - x \leq e^{-x}$$

for all real numbers x.

5.10 Let $(S, \Pr)$ be a probability space and let B be an event with $\Pr(B) > 0$. Define the function $\Pr' : S \to \mathbb{R}$ by

$$\Pr'(\omega) = \begin{cases} \frac{\Pr(\omega)}{\Pr(B)} & \text{if } \omega \in B, \\ 0 & \text{if } \omega \notin B. \end{cases}$$

- Prove that $\Pr'$ is a probability function on S according to Definition 5.2.2.
- Prove that for any event A,

$$\Pr'(A) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

5.11 You roll two dice D_1 and D_2. Let A be the event "$D_1 + D_2 = 7$" and let B be the event "$D_1 = 4$". Determine the conditional probabilities $\Pr(A \mid B)$ and $\Pr(B \mid A)$.

5.12 A hand of 13 cards is chosen uniformly at random from a standard deck of 52 cards. Define the following events:

- $A =$ "the hand has at least one ace",
- $B =$ "the hand has at least two aces",
- $C =$ "the hand has the ace of spades".

Determine the conditional probabilities $\Pr(A \mid B)$, $\Pr(B \mid A)$, and $\Pr(B \mid C)$.

5.13 We take a uniformly random permutation of a standard deck of 52 cards, so that each permutation has a probability of $1/52!$. Define the following events:

- $A =$ "the top card is an ace",
- $B =$ "the bottom card is the ace of spades",
- $C =$ "the bottom card is the queen of spades".

Determine the conditional probabilities $\Pr(A \mid B)$ and $\Pr(A \mid C)$.

5.14 Consider two dice, each one having one face showing the letter a, two faces showing the letter b, and the remaining three faces showing the letter c. You roll each die once, independently of the other die.

- What is the sample space?
- Define the events

  $A =$ "at least one of the two dice shows the letter b on its top face"

  and

  $B =$ "both dice show the same letter on their top faces".

  Determine $\Pr(A)$, $\Pr(B)$, and $\Pr(A \mid B)$.

5.15 You flip a fair coin, independently, three times. Define the events

$A =$ "the first flip results in heads",

and

$B =$ "the coin comes up heads exactly once".

Determine $\Pr(A \mid B)$ and $\Pr(B \mid A)$.

5.16 According to Statistics Canada, a random person in Canada has

- a probability of 4/5 to live to at least 70 years old and
- a probability of 1/2 to live to at least 80 years old.

John (a random person in Canada) has just celebrated his 70-th birthday. What is the probability that John will celebrate his 80-th birthday?

5.17 Let A and B be events with $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$. Use the definition of conditional probability to prove Bayes' Theorem:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)}.$$

5.18 Medical doctors have developed a test for detecting disease X.

- The test is 98% effective on people who have X: If a person has X, then with probability 0.98, the test says that the person indeed has X.
- The test gives a false reading for 3% of the population without the disease: If a person does not have X, then with probability 0.03, the test says that the person does have X.
- It is known that 0.1% of the population has X.

Assume we choose a person uniformly at random from the population and test this person for disease X.

- Determine the probability that the test says that the person has X.
- Assume the test says that the person has X. Use Exercise 5.17 to determine the probability that the person indeed has X.

5.19 Consider three events A, B, and C in a sample space S, and assume that $\Pr(B \cap C) \neq 0$ and $\Pr(C) \neq 0$. Prove that

$$\Pr(A \cap B \cap C) = \Pr(A \mid B \cap C) \cdot \Pr(B \mid C) \cdot \Pr(C).$$

5.20 You have a fair die and do the following experiment:

- Roll the die once; let x be the outcome.

- Roll the die x times (independently); let y be the smallest outcome of these x rolls.

- Roll the die y times (independently); let z be the largest outcome of these y rolls.

Use Exercise 5.19 to determine

$$\Pr(x = 1 \text{ and } y = 2 \text{ and } z = 3).$$

5.21 A hand of 5 cards is chosen uniformly at random from a standard deck of 52 cards. Define the event

$$A = \text{ “the hand has at least one ace”.}$$

- Explain what is wrong with the following argument:

> We are going to determine $\Pr(A)$. Event A states that the hand has at least one ace. By symmetry, we may assume that A is the event that the hand has the ace of spades. Since there are $\binom{52}{5}$ hands of five cards and exactly $\binom{51}{4}$ of them contain the ace of spades, it follows that
>
> $$\Pr(A) = \frac{\binom{51}{4}}{\binom{52}{5}} = \frac{5}{52}.$$

- Explain what is wrong with the following argument:

We are going to determine $\Pr(A)$ using the Law of Total Probability (Theorem 5.9.1). For each $x \in \{\spadesuit, \heartsuit, \clubsuit, \diamondsuit\}$, define the event

$$B_x = \text{ “the hand has the ace of suit } x\text{”}.$$

We observe that

$$\Pr(B_x) = \frac{\binom{51}{4}}{\binom{52}{5}} = \frac{5}{52}.$$

We next observe that

$$\Pr(A \mid B_x) = 1,$$

because if event B_x occurs, then event A also occurs. Thus, using the Law of Total Probability, we get

$$\begin{aligned}
\Pr(A) &= \sum_x \Pr(A \mid B_x) \cdot \Pr(B_x) \\
&= \sum_x 1 \cdot \Pr(B_x) \\
&= \sum_x \frac{5}{52} \\
&= 4 \cdot \frac{5}{52} \\
&= \frac{5}{13}.
\end{aligned}$$

- Determine the value of $\Pr(A)$.

5.22 You are doing two projects P and Q. The probability that project P is successful is equal to $2/3$ and the probability that project Q is successful is equal to $4/5$. Whether or not these two projects are successful are independent of each other. What is the probability that both P and Q are not successful?

5.23 Consider two independent events A and B in a sample space S. Assume that A and B are disjoint, i.e., $A \cap B = \emptyset$. What can you say about $\Pr(A)$ and $\Pr(B)$?

5.24 You flip three fair coins independently of each other. Let A be the event "at least two flips in a row are heads" and let B be the event "the number of heads is even". (Note that zero is even.) Are A and B independent?

5.25 You flip three fair coins independently of each other. Let A be the event "there is at most one tails" and let B be the event "not all flips are identical". Are A and B independent?

5.26 You flip two fair coins independently of each other. Define the following events:

$$A = \text{ "the number of heads is odd"},$$

$$B = \text{ "the first coin comes up heads"},$$

and

$$C = \text{ "the second coin comes up heads"}.$$

- Are the events A and B independent?
- Are the events A and C independent?
- Are the events B and C independent?
- Are the events A, B, and C pairwise independent?
- Are the events A, B, and C mutually independent?

5.27 You are given a tetrahedron, which is a die with four faces. Each of these faces has one of the bitstrings 110, 101, 011, and 000 written on it. Different faces have different bitstrings.

We roll the tetrahedron so that each face is at the bottom with equal probability 1/4. For $k = 1, 2, 3$, define the event

$$A_k = \text{ "the bitstring written on the bottom face has 0 at position } k\text{"}.$$

For example, if the bitstring at the bottom face is 101, then A_1 is false, A_2 is true, and A_3 is false.

- Are the events A_1 and A_2 independent?
- Are the events A_1 and A_3 independent?

- Are the events A_2 and A_3 independent?
- Are the events A_1, A_2, A_3 pairwise independent?
- Are the events A_1, A_2, A_3 mutually independent?

5.28 In a group of 100 children, 34 are boys and 66 are girls. You are given the following information about the girls:

- Each girl has green eyes or is blond or is left-handed.
- 20 of the girls have green eyes.
- 40 of the girls are blond.
- 50 of the girls are left-handed.
- 10 of the girls have green eyes and are blond.
- 14 of the girls have green eyes and are left-handed.
- 4 of the girls have green eyes, are blond, and are left-handed.

We choose one of these 100 children uniformly at random. Define the events

$$G = \text{ “the kid chosen is a girl with green eyes”},$$

$$B = \text{ “the kid chosen is a blond girl”},$$

and

$$L = \text{ “the kid chosen is a left-handed girl”}.$$

- Are the events G and B independent?
- Are the events G and L independent?
- Are the events B and L independent?
- Verify whether or not the following equation holds:

$$\Pr(G \wedge B \wedge L) = \Pr(G) \cdot \Pr(B) \cdot \Pr(L).$$

5.29 Annie, Boris, and Charlie write an exam that consists of only one question: *What is* 26 *times* 26*?* Calculators are not allowed during the exam. Both Annie and Boris are pretty clever and each of them gives the correct answer with probability 9/10. Charlie has trouble with two-digit numbers and gives the correct answer with probability 6/10.

- Assume that the three students do not cheat, i.e., each student answers the question independently of the other two students. Determine the probability that at least two of them give the correct answer.
- Assume that Annie and Boris do not cheat, but Charlie copies Annie's answer. Determine the probability that at least two of them give the correct answer.

Hint: The answer to the second part is smaller than the answer to the first part.

5.30 Prove that for any real number $x \neq 1$ and any integer $N \geq 0$,

$$\sum_{n=0}^{N} x^n = \frac{1 - x^{N+1}}{1 - x}.$$

5.31 Use the following argumentation to convince yourself that

$$\sum_{n=0}^{\infty} 1/2^n = 2.$$

Take the interval $I = [0, 2)$ of length 2 on the real line and, for each $n \geq 0$, an interval I_n of length $1/2^n$. It is possible to place all intervals I_n with $n \geq 0$ in I such that

- no two intervals I_n and I_m, with $m \neq n$, overlap and
- all intervals I_n with $n \geq 0$ completely cover the interval I.

5.32 Two players P and Q take turns rolling two fair and independent dice. The first player who gets a sum of seven wins the game. Determine the probability that player P wins the game.

5.33 By flipping a fair coin repeatedly and independently, we obtain a sequence of H's and T's. We stop flipping the coin as soon as the sequence contains either HH or TH.

Two players play a game, in which Player 1 wins if the last two symbols in the sequence are HH. Otherwise, the last two symbols in the sequence are TH, in which case Player 2 wins. Define the events

$$A = \text{ “Player 1 wins”}$$

and

$$B = \text{ “Player 2 wins.”}$$

Determine $\Pr(A)$ and $\Pr(B)$.

5.34 We flip a fair coin repeatedly and independently, and stop as soon as we see one of the two sequences HTT and HHT. Let A be the event that the process stops because HTT is seen.

- Prove that the event A is given by the set

  $$\{T^m(HT)^nHTT : m \geq 0, n \geq 0\}.$$

  In other words, event A holds if and only if the sequence of coin flips is equal to $T^m(HT)^nHTT$ for some $m \geq 0$ and $n \geq 0$.

- Prove that $\Pr(A) = 1/3$.

5.35 For $i \in \{1, 2\}$, consider the game G_i, in which two players P_1 and P_2 take turns flipping, independently, a fair coin, where P_i starts. The game ends as soon as heads comes up. The player who flips heads first is the winner of the game G_i. For $j \in \{1, 2\}$, define the event

$$B_{ij} = \text{“}P_j \text{ wins the game } G_i\text{”}.$$

In Section 5.14.2, we have seen that

$$\Pr(B_{11}) = \Pr(B_{22}) = 2/3 \tag{5.6}$$

and

$$\Pr(B_{12}) = \Pr(B_{21}) = 1/3. \tag{5.7}$$

Consider the game G, in which P_1 and P_2 take turns flipping, independently, a fair coin, where P_1 starts. The game ends as soon as a second heads comes up. The player who flips the second heads wins the game. Define the event

$$A = \text{“}P_1 \text{ wins the game } G\text{”}.$$

In Section 5.14.3, we used an infinite series to show that

$$\Pr(A) = 4/9. \tag{5.8}$$

Use the Law of Total Probability (Theorem 5.9.1) to give an alternative proof of (5.8). You are allowed to use (5.6) and (5.7).

5.36 You would like to generate a uniformly random bit, i.e., with probability 1/2, this bit is 0, and with probability 1/2, it is 1. You find a coin in your pocket, but you are not sure if it is a fair coin: It comes up heads (H) with probability p and tails (T) with probability $1-p$, for some real number p that is unknown to you. In particular, you do not know if $p = 1/2$. In this exercise, you will show that this coin can nevertheless be used to generate a uniformly random bit.

Consider the following recursive algorithm GetRandomBit, which does not take any input:

Algorithm GetRandomBit**:**

```
// all coin flips made are mutually independent
flip the coin twice;
if the result is HT
then return 0
else if the result is TH
     then return 1
     else GetRandomBit
     endif
endif
```

- The sample space S is the set of all sequences of coin flips that can occur when running algorithm GetRandomBit. Determine this sample space S.
- Prove that algorithm GetRandomBit returns a uniformly random bit.

Chapter 6

Random Variables and Expectation

6.1 Random Variables

We have already seen random variables in Chapter 5, even though we did not use that term there. For example, in Section 5.2.1, we rolled a die twice and were interested in the sum of the results of these two rolls. In other words, we did an "experiment" (rolling a die twice) and asked for a function of the outcome (the sum of the results of the two rolls).

Definition 6.1.1 Let S be a sample space. A *random variable* is a function $X : S \to \mathbb{R}$.

In the example given above, the sample space is

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$$

and the random variable is the function $X : S \to \mathbb{R}$ defined by

$$X(i, j) = i + j$$

for all (i, j) in S.

6.1.1 Flipping Three Coins

Assume we flip three coins. The sample space is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\},$$

where, e.g., TTH indicates that the first two coins come up tails and the third coin comes up heads.

Let $X : S \to \mathbb{R}$ be the random variable that maps any outcome (i.e., any element of S) to the number of heads in the outcome. Thus,

$$\begin{aligned} X(HHH) &= 3, \\ X(HHT) &= 2, \\ X(HTH) &= 2, \\ X(HTT) &= 1, \\ X(THH) &= 2, \\ X(THT) &= 1, \\ X(TTH) &= 1, \\ X(TTT) &= 0. \end{aligned}$$

If we define the random variable Y to be the function $Y : S \to \mathbb{R}$ that

- maps an outcome to 1 if all three coins come up heads or all three coins come up tails, and
- maps an outcome to 0 in all other cases,

then we have

$$\begin{aligned} Y(HHH) &= 1, \\ Y(HHT) &= 0, \\ Y(HTH) &= 0, \\ Y(HTT) &= 0, \\ Y(THH) &= 0, \\ Y(THT) &= 0, \\ Y(TTH) &= 0, \\ Y(TTT) &= 1. \end{aligned}$$

Since a random variable is a function $X : S \to \mathbb{R}$, it maps any outcome ω to a real number $X(\omega)$. Usually, we just write X instead of $X(\omega)$. Thus, for any outcome in the sample space S, we denote the value of the random variable, for this outcome, by X. In the example above, we flip three coins and write

$$X = \text{ the number of heads}$$

and

$$Y = \begin{cases} 1 & \text{if all three coins come up heads or all three coins come up tails,} \\ 0 & \text{otherwise.} \end{cases}$$

6.1.2 Random Variables and Events

Random variables give rise to events in a natural way. In the three-coin example, "$X = 0$" corresponds to the event $\{TTT\}$, whereas "$X = 2$" corresponds to the event $\{HHT, HTH, THH\}$. The table below gives some values of the random variables X and Y, together with the corresponding events.

value	event
$X = 0$	$\{TTT\}$
$X = 1$	$\{HTT, THT, TTH\}$
$X = 2$	$\{HHT, HTH, THH\}$
$X = 3$	$\{HHH\}$
$X = 4$	$\emptyset$
$Y = 0$	$\{HHT, HTH, HTT, THH, THT, TTH\}$
$Y = 1$	$\{HHH, TTT\}$
$Y = 2$	$\emptyset$

Thus, the event "$X = x$" corresponds to the set of all outcomes that are mapped, by the function X, to the value x:

Definition 6.1.2 Let S be a sample space and let $X : S \to \mathbb{R}$ be a random variable. For any real number x, we define "$X = x$" to be the event $\{\omega \in S : X(\omega) = x\}$.

Let us return to the example in which we flip three coins. Assume that the coins are fair and the three flips are mutually independent. Consider again the corresponding random variables X and Y. It should be clear how we determine, for example, the probability that X is equal to 0, which we will write as $\Pr(X = 0)$. Using our interpretation of "$X = 0$" as being the event $\{TTT\}$, we get

$$\begin{aligned} \Pr(X = 0) &= \Pr(TTT) \\ &= 1/8. \end{aligned}$$

Similarly, we get

$$\begin{aligned}
\Pr(X = 1) &= \Pr(\{HTT, THT, TTH\}) \\
&= 3/8, \\
\Pr(X = 2) &= \Pr(\{HHT, HTH, THH\}) \\
&= 3/8, \\
\Pr(X = 3) &= \Pr(\{HHH\}) \\
&= 1/8, \\
\Pr(X = 4) &= \Pr(\emptyset) \\
&= 0, \\
\Pr(Y = 0) &= \Pr(\{HHT, HTH, HTT, THH, THT, TTH\}) \\
&= 6/8 \\
&= 3/4, \\
\Pr(Y = 1) &= \Pr(\{HHH, TTT\}) \\
&= 2/8 \\
&= 1/4, \\
\Pr(Y = 2) &= \Pr(\emptyset) \\
&= 0.
\end{aligned}$$

Consider an arbitrary probability space $(S, \Pr)$ and let $X : S \to \mathbb{R}$ be a random variable. Using (5.1) and Definition 6.1.2, the probability of the event "$X = x$", i.e., the probability that X is equal to x, is equal to

$$\begin{aligned}
\Pr(X = x) &= \Pr(\{\omega \in S : X(\omega) = x\}) \\
&= \sum_{\omega : X(\omega) = x} \Pr(\omega).
\end{aligned}$$

We have interpreted "$X = x$" as being an event. We extend this to more general statements involving X. For example, "$X \geq x$" denotes the event

$$\{\omega \in S : X(\omega) \geq x\}.$$

For our three-coin example, the random variable X can take each of the values 0, 1, 2, and 3 with a positive probability. As a result, "$X \geq 2$"

denotes the event "$X = 2$ or $X = 3$", and we have

$$\begin{aligned} \Pr(X \geq 2) &= \Pr(X = 2 \vee X = 3) \\ &= \Pr(X = 2) + \Pr(X = 3) \\ &= 3/8 + 1/8 \\ &= 1/2. \end{aligned}$$

6.2 Independent Random Variables

In Section 5.10, we have defined the notion of two events being independent. The following definition extends this to random variables.

Definition 6.2.1 Let $(S, \Pr)$ be a probability space and let X and Y be two random variables. We say that X and Y are *independent* if for all real numbers x and y, the events "$X = x$" and "$Y = y$" are independent, i.e.,

$$\Pr(X = x \wedge Y = y) = \Pr(X = x) \cdot \Pr(Y = y).$$

Assume we flip three fair coins independently and, as in Section 6.1.1, consider the random variables

$$X = \text{ the number of heads}$$

and

$$Y = \begin{cases} 1 & \text{if all three coins come up heads or all three coins come up tails,} \\ 0 & \text{otherwise.} \end{cases}$$

Are these two random variables independent? Observe the following: If $Y = 1$, then $X = 0$ or $X = 3$. In other words, if we are given some information about the random variable Y (in this case, $Y = 1$), then the random variable X cannot take, for example, the value 2. Based on this, we take $x = 2$ and $y = 1$ in Definition 6.2.1. Since the event "$X = 2 \wedge Y = 1$" is equal to $\emptyset$, we have

$$\Pr(X = 2 \wedge Y = 1) = \Pr(\emptyset) = 0.$$

On the other hand, we have seen in Section 6.1.2 that $\Pr(X = 2) = 3/8$ and $\Pr(Y = 1) = 1/4$. It follows that

$$\Pr(X = 2 \wedge Y = 1) \neq \Pr(X = 2) \cdot \Pr(Y = 1)$$

and, therefore, the random variables X and Y are not independent.

Now consider the random variable

$$Z = \begin{cases} 1 & \text{if the first coin comes up heads,} \\ 0 & \text{if the first coin comes up tails.} \end{cases}$$

We claim that the random variables Y and Z are independent. To verify this, we have to show that for all real numbers y and z,

$$\Pr(Y = y \wedge Z = z) = \Pr(Y = y) \cdot \Pr(Z = z). \tag{6.1}$$

Recall from Section 6.1.2 that $\Pr(Y = 1) = 1/4$ and $\Pr(Y = 0) = 3/4$. Since the coin flips are independent, we have $\Pr(Z = 1) = 1/2$ and $\Pr(Z = 0) = 1/2$. Furthermore,

$$\begin{aligned} \Pr(Y = 1 \wedge Z = 1) &= \Pr(HHH) \\ &= 1/8, \\ \Pr(Y = 1 \wedge Z = 0) &= \Pr(TTT) \\ &= 1/8, \\ \Pr(Y = 0 \wedge Z = 1) &= \Pr(HHT, HTH, HTT) \\ &= 3/8, \\ \Pr(Y = 0 \wedge Z = 0) &= \Pr(THH, THT, TTH) \\ &= 3/8. \end{aligned}$$

It follows that

$$\Pr(Y = 1 \wedge Z = 1) = \Pr(Y = 1) \cdot \Pr(Z = 1),$$

$$\Pr(Y = 1 \wedge Z = 0) = \Pr(Y = 1) \cdot \Pr(Z = 0),$$

$$\Pr(Y = 0 \wedge Z = 1) = \Pr(Y = 0) \cdot \Pr(Z = 1),$$

and

$$\Pr(Y = 0 \wedge Z = 0) = \Pr(Y = 0) \cdot \Pr(Z = 0).$$

Thus, (6.1) holds if $(y, z) \in \{(1, 1), (1, 0), (0, 1), (0, 0)\}$. For any other pair (y, z), such as $(y, z) = (3, 5)$ or $(y, z) = (1, 2)$, at least one of the events "$Y = y$" and "$Z = z$" is the empty set, i.e., cannot occur. Therefore, for such pairs, we have

$$\Pr(Y = y \wedge Z = z) = 0 = \Pr(Y = y) \cdot \Pr(Z = z).$$

Thus, we have indeed verified that (6.1) holds for all real numbers y and z. As a result, we have shown that the random variables Y and Z are independent.

Are the random variables X and Z independent? If $X = 0$, then all three coins come up tails and, therefore, $Z = 0$. Thus,

$$\Pr(X = 0 \wedge Z = 1) = \Pr(\emptyset) = 0,$$

whereas

$$\Pr(X = 0) \cdot \Pr(Z = 1) = 1/8 \cdot 1/2 \neq 0.$$

As a result, the random variables X and Z are not independent.

We have defined the notion of two random variables being independent. As in Definition 5.10.3, there are two ways to generalize this to sequences of random variables:

Definition 6.2.2 Let $(S, \Pr)$ be a probability space, let $n \geq 2$, and let $X_1, X_2, \ldots, X_n$ be a sequence of random variables.

1. We say that this sequence is *pairwise independent* if for all real numbers $x_1, x_2, \ldots, x_n$, the sequence "$X_1 = x_1$", "$X_2 = x_2$", ..., "$X_n = x_n$" of events is pairwise independent.

2. We say that this sequence is *mutually independent* if for all real numbers $x_1, x_2, \ldots, x_n$, the sequence "$X_1 = x_1$", "$X_2 = x_2$", ..., "$X_n = x_n$" of events is mutually independent.

6.3 Distribution Functions

Consider a random variable X. In Section 6.1.2, we have defined $\Pr(X = x)$, i.e., the probability of the event "$X = x$", to be

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

This defines a function that maps any real number x to the real number $\Pr(X = x)$. This function is called the distribution function of the random variable X:

Definition 6.3.1 Let $(S, \Pr)$ be a probability space and let $X : S \to \mathbb{R}$ be a random variable. The *distribution function* of X is the function $D : \mathbb{R} \to \mathbb{R}$ defined by

$$D(x) = \Pr(X = x)$$

for all $x \in \mathbb{R}$.

For example, assume we roll two independent fair dice, so that the sample space is

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\},$$

where i is the result of the first roll and j is the result of the second roll. Each outcome (i, j) in S has the same probability of $1/36$.

Let X be the random variable whose value is equal to the sum of the results of the two rolls. The matrix below gives all possible values of X. The leftmost column gives the result of the first roll, the top row gives the result of the second roll, and each other entry is the corresponding value of X.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

As can be seen from this matrix, the random variable X can take any value in $\{2, 3, 4, \ldots, 12\}$. The distribution function D of X is given by

$$\begin{aligned}
D(2) &= \Pr(X = 2) &&= 1/36, \\
D(3) &= \Pr(X = 3) &&= 2/36, \\
D(4) &= \Pr(X = 4) &&= 3/36, \\
D(5) &= \Pr(X = 5) &&= 4/36, \\
D(6) &= \Pr(X = 6) &&= 5/36, \\
D(7) &= \Pr(X = 7) &&= 6/36, \\
D(8) &= \Pr(X = 8) &&= 5/36, \\
D(9) &= \Pr(X = 9) &&= 4/36, \\
D(10) &= \Pr(X = 10) &&= 3/36, \\
D(11) &= \Pr(X = 11) &&= 2/36, \\
D(12) &= \Pr(X = 12) &&= 1/36,
\end{aligned}$$

whereas for all $x \notin \{2, 3, 4, \ldots, 12\}$,

$$D(x) = \Pr(X = x) = 0.$$

In Sections 6.6 and 6.7, we will see other examples of distribution functions.

6.4 Expected Values

Consider the probability space $(S, \Pr)$ with sample space $S = \{1, 2, 3\}$ and probability function Pr defined by $\Pr(1) = 4/5$, $\Pr(2) = 1/10$, and $\Pr(3) = 1/10$. Assume we choose an element in S according to this probability function. Let X be the random variable whose value is equal to the element in S that is chosen. Thus, as a function $X : S \to \mathbb{R}$, we have $X(1) = 1$, $X(2) = 2$, and $X(3) = 3$.

The "expected value" of X is the value of X that we observe "on average". How should we define this? Since X has a much higher probability to take the value 1 than the other two values 2 and 3, the value 1 should get a larger "weight" in the expected value of X. Based on this, it is natural to define the expected value of X to be

$$1 \cdot \Pr(1) + 2 \cdot \Pr(2) + 3 \cdot \Pr(3) = 1 \cdot \frac{4}{5} + 2 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} = \frac{13}{10}.$$

Definition 6.4.1 Let $(S, \Pr)$ be a probability space and let $X : S \to \mathbb{R}$ be a random variable. The *expected value* of X is defined to be

$$\mathbb{E}(X) = \sum_{\omega \in S} X(\omega) \cdot \Pr(\omega),$$

provided this summation converges absolutely[1].

6.4.1 Some Examples

Flipping a coin: Assume we flip a fair coin, in which case the sample space is $S = \{H, T\}$ and $\Pr(H) = \Pr(T) = 1/2$. Define the random variable

[1]The series $\sum_{n=0}^{\infty} a_n$ converges absolutely if the series $\sum_{n=0}^{\infty} |a_n|$ converges. If a series converges absolutely, then we can change the order of summation without changing the value of the series.

X to have value

$$X = \begin{cases} 1 & \text{if the coin comes up heads,} \\ 0 & \text{if the coin comes up tails.} \end{cases}$$

Thus, as a function $X : S \to \mathbb{R}$, we have $X(H) = 1$ and $X(T) = 0$. The expected value $\mathbb{E}(X)$ of X is equal to

$$\begin{aligned} \mathbb{E}(X) &= X(H) \cdot \Pr(H) + X(T) \cdot \Pr(T) \\ &= 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \\ &= \frac{1}{2}. \end{aligned}$$

This example shows that the term "expected value" is a bit misleading: $\mathbb{E}(X)$ is *not* the value that we expect to observe, because the value of X is *never* equal to its expected value.

Rolling a die: Assume we roll a fair die. Define the random variable X to be the value of the result. Then, X takes each of the values in $\{1, 2, 3, 4, 5, 6\}$ with equal probability $1/6$, and we get

$$\begin{aligned} \mathbb{E}(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2}. \end{aligned}$$

Now define the random variable Y to be equal to one divided by the result of the die. In other words, $Y = 1/X$. This random variable takes each of the values in $\{1, 1/2, 1/3, 1/4, 1/5, 1/6\}$ with equal probability $1/6$, and we get

$$\begin{aligned} \mathbb{E}(Y) &= 1 \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} \\ &= \frac{49}{120}. \end{aligned}$$

Note that $\mathbb{E}(Y) \neq 1/\mathbb{E}(X)$. Thus, this example shows that, in general, $\mathbb{E}(1/X) \neq 1/\mathbb{E}(X)$.

Rolling two dice: Assume we roll two independent fair dice, in which case the sample space is

$$S = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\},$$

where i is the result of the first roll and j is the result of the second roll. Each outcome (i, j) in S has the same probability of $1/36$.

Let X be the random variable whose value is equal to the sum of the results of the two rolls. As a function $X : S \to \mathbb{R}$, we have $X(i, j) = i+j$. The matrix below gives all possible values of X. The leftmost column indicates the result of the first roll, the top row indicates the result of the second roll, and each other entry is the corresponding value of X.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The expected value $\mathbb{E}(X)$ of X is equal to

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{(i,j)\in S} X(i, j) \cdot \Pr(i, j) \\
&= \sum_{(i,j)\in S} (i + j) \cdot \frac{1}{36} \\
&= \frac{1}{36} \sum_{(i,j)\in S} (i + j) \\
&= \frac{1}{36} \cdot \text{the sum of all 36 entries in the matrix} \\
&= \frac{1}{36} \cdot 252 \\
&= 7.
\end{aligned}$$

6.4.2 An Alternative Formula for the Expected Value

In the last example, we used Definition 6.4.1 to compute the expected value $\mathbb{E}(X)$ of the random variable X that was defined to be the sum of the results

when rolling two dice. This was a painful way to compute $\mathbb{E}(X)$, because we added all 36 entries in the matrix. There is a slightly easier way to determine $\mathbb{E}(X)$: By looking at the matrix, we see, for example, that there is 1 way for the event "$X = 2$" to occur, whereas there are 2 ways for the event "$X = 3$" to occur. The table below lists the number of ways that each possible event can occur.

event	number of ways
$X = 2$	1
$X = 3$	2
$X = 4$	3
$X = 5$	4
$X = 6$	5
$X = 7$	6
$X = 8$	5
$X = 9$	4
$X = 10$	3
$X = 11$	2
$X = 12$	1

Based on this, we get

$$\begin{aligned}\mathbb{E}(X) &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + \\ &= 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\ &= 7.\end{aligned}$$

Even though this is still quite painful, less computation is needed. What we have done is the following: In the definition of $\mathbb{E}(X)$, i.e.,

$$\mathbb{E}(X) = \sum_{(i,j) \in S} X(i,j) \cdot \Pr(i,j),$$

we rearranged the terms in the summation. That is, instead of taking the sum over all elements (i, j) in S, we

- grouped together all outcomes (i, j) for which $X(i, j) = i + j$ has the same value, say, k,

- multiplied this common value k by the probability that X is equal to k,
- and took the sum of the resulting products over all possible values of k.

This resulted in

$$\mathbb{E}(X) = \sum_{k=2}^{12} k \cdot \Pr(X = k).$$

The following lemma states that this can be done for any random variable.

Lemma 6.4.2 *Let $(S, \Pr)$ be a probability space and let $X : S \to \mathbb{R}$ be a random variable. The expected value of X is equal to*

$$\mathbb{E}(X) = \sum_{x} x \cdot \Pr(X = x).$$

Proof. Recall that the event "$X = x$" corresponds to the subset

$$A_x = \{\omega \in S : X(\omega) = x\}$$

of the sample space S. We have

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\omega \in S} X(\omega) \cdot \Pr(\omega) \\
&= \sum_{x} \sum_{\omega : X(\omega) = x} X(\omega) \cdot \Pr(\omega) \\
&= \sum_{x} \sum_{\omega : X(\omega) = x} x \cdot \Pr(\omega) \\
&= \sum_{x} \sum_{\omega \in A_x} x \cdot \Pr(\omega) \\
&= \sum_{x} x \sum_{\omega \in A_x} \Pr(\omega) \\
&= \sum_{x} x \cdot \Pr\left(A_x\right) \\
&= \sum_{x} x \cdot \Pr(X = x).
\end{aligned}$$

■

When computing the expected value of a random variable X, it is usually easier to use Lemma 6.4.2 than Definition 6.4.1. To use Lemma 6.4.2, you have to do the following:

- Determine the values x that X can take.
- For each such value x, determine $\Pr(X = x)$.
- Compute the sum of all values $x \cdot \Pr(X = x)$.

6.5 Linearity of Expectation

We now come to one of the most useful tools for determining expected values:

Theorem 6.5.1 *Let $(S, \Pr)$ be a probability space. For any two random variables X and Y on S, and for any two real numbers a and b,*

$$\mathbb{E}(aX + bY) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).$$

Proof. Recall that both X and Y are functions from S to $\mathbb{R}$. Define the random variable Z to be $Z = aX + bY$. That is, as a function $Z : S \to \mathbb{R}$, Z is defined by

$$Z(\omega) = a \cdot X(\omega) + b \cdot Y(\omega)$$

for all ω in S. Using Definition 6.4.1, we get

$$\begin{aligned}
\mathbb{E}(Z) &= \sum_{\omega \in S} Z(\omega) \cdot \Pr(\omega) \\
&= \sum_{\omega \in S} (a \cdot X(\omega) + b \cdot Y(\omega)) \cdot \Pr(\omega) \\
&= a \sum_{\omega \in S} X(\omega) \cdot \Pr(\omega) + b \sum_{\omega \in S} Y(\omega) \cdot \Pr(\omega) \\
&= a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).
\end{aligned}$$

■

Let us return to the example in which we roll two independent fair dice and define the random variable X to be the sum of the results of the two

rolls. We have seen two ways to compute the expected value $\mathbb{E}(X)$ of X. We now present a third way, which is the easiest one: We define two random variables

$$Y = \text{ the result of the first die}$$

and

$$Z = \text{ the result of the second die.}$$

We have already seen that

$$\mathbb{E}(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

Of course, by the same computation, we have

$$\mathbb{E}(Z) = \frac{7}{2}.$$

Observe that

$$X = Y + Z.$$

Then, by the Linearity of Expectation (i.e., Theorem 6.5.1), we have

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(Y + Z) \\ &= \mathbb{E}(Y) + \mathbb{E}(Z) \\ &= \frac{7}{2} + \frac{7}{2} \\ &= 7. \end{aligned}$$

We have stated the Linearity of Expectation for two random variables. The proof of Theorem 6.5.1 can easily be generalized to any finite sequence of random variables:

Theorem 6.5.2 *Let $(S, \Pr)$ be a probability space, let $n \geq 2$ be an integer, let $X_1, X_2, \ldots, X_n$ be a sequence of random variables, and let $a_1, a_2, \ldots, a_n$ be a sequence of real numbers. Then,*

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \cdot \mathbb{E}\left(X_i\right).$$

The following theorem states that the Linearity of Expectation also holds for infinite sequences of random variables:

Theorem 6.5.3 *Let $(S, \Pr)$ be a probability space and let $X_1, X_2, \ldots$ be an infinite sequence of random variables such that the infinite series*

$$\sum_{i=1}^{\infty} \mathbb{E}\left(|X_i|\right)$$

converges. Then,

$$\mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} \mathbb{E}\left(X_i\right).$$

Proof. Define the random variable X to be

$$X = \sum_{i=1}^{\infty} X_i.$$

The derivation below uses Definition 6.4.1 and the assumption that the infinite series $\sum_{i=1}^{\infty} \mathbb{E}\left(|X_i|\right)$ converges, which allows us to change the order of summation without changing the value of the series:

$$\begin{aligned}
\sum_{i=1}^{\infty} \mathbb{E}\left(X_i\right) &= \sum_{i=1}^{\infty} \sum_{\omega \in S} X_i(\omega) \cdot \Pr(\omega) \\
&= \sum_{\omega \in S} \sum_{i=1}^{\infty} X_i(\omega) \cdot \Pr(\omega) \\
&= \sum_{\omega \in S} \Pr(\omega) \sum_{i=1}^{\infty} X_i(\omega) \\
&= \sum_{\omega \in S} \Pr(\omega) \cdot X(\omega) \\
&= \mathbb{E}(X) \\
&= \mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right).
\end{aligned}$$

■

6.6 The Geometric Distribution

Let p be a real number with $0 < p < 1$ and consider an experiment that is *successful* with probability p and *fails* with probability $1-p$. We repeat this experiment independently until it is successful for the first time. What is the expected number of times that we perform the experiment?

We model this problem in the following way: Assume we have a coin that comes up heads with probability p and, thus, comes up tails with probability $1-p$. We flip this coin repeatedly and independently until it comes up heads for the first time. (We have seen this process in Section 5.14 for the case when $p = 1/2$.) Define the random variable X to be the number of times that we flip the coin; this includes the last coin flip, which resulted in heads. We want to determine the expected value $\mathbb{E}(X)$ of X.

The sample space is given by

$$S = \{T^{k-1}H : k \geq 1\}.$$

Since the coin flips are independent, the outcome $T^{k-1}H$ has a probability of $(1-p)^{k-1}p = p(1-p)^{k-1}$, i.e.,

$$\Pr\left(T^{k-1}H\right) = p(1-p)^{k-1}.$$

Let us first verify that all probabilities add up to 1: Using Lemma 5.14.2, we have

$$\begin{aligned}
\sum_{k=1}^{\infty} \Pr\left(T^{k-1}H\right) &= \sum_{k=1}^{\infty} p(1-p)^{k-1} \\
&= p\sum_{k=1}^{\infty} (1-p)^{k-1} \\
&= p\sum_{\ell=0}^{\infty} (1-p)^{\ell} \\
&= p \cdot \frac{1}{1-(1-p)} \\
&= 1.
\end{aligned}$$

6.6.1 Determining the Expected Value

We are going to use Lemma 6.4.2 to determine the expected value $\mathbb{E}(X)$. We first observe that X can take any value in $\{1, 2, 3, \ldots\}$. For any integer

$k \geq 1$, $X = k$ if and only if the coin flips give the sequence $T^{k-1}H$. It follows that

$$\Pr(X = k) = \Pr\left(T^{k-1}H\right) = p(1-p)^{k-1}. \tag{6.2}$$

By Lemma 6.4.2, we have

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{\infty} k \cdot \Pr(X = k) \\
&= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \\
&= p\sum_{k=1}^{\infty} k(1-p)^{k-1}.
\end{aligned}$$

How do we determine the infinite series on the right-hand side?

According to Lemma 5.14.2, we have

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},$$

for any real number x with $-1 < x < 1$. Both sides of this equation are functions of x and these two functions are equal to each other. If we differentiate both sides, we get two derivatives that are also equal to each other:

$$\sum_{k=0}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

Since for $k = 0$, the term kx^{k-1} is equal to 0, we have

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

If we take $x = 1 - p$, we get

$$\begin{aligned}
\mathbb{E}(X) &= p\sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= p \cdot \frac{1}{(1-(1-p))^2} \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}.
\end{aligned}$$

In Section 6.3, we have defined the distribution function of a random variable. The distribution function of the coin-flipping random variable X is given by (6.2). This function is called a geometric distribution:

Definition 6.6.1 Let p be a real number with $0 < p < 1$. A random variable X has a *geometric distribution with parameter* p, if its distribution function satisfies

$$\Pr(X = k) = p(1-p)^{k-1}$$

for any integer $k \geq 1$.

Our calculation that led to the value of $\mathbb{E}(X)$ proves the following theorem:

Theorem 6.6.2 *Let p be a real number with $0 < p < 1$ and let X be a random variable that has a geometric distribution with parameter p. Then*

$$\mathbb{E}(X) = 1/p.$$

For example, if we flip a fair coin (in which case $p = 1/2$) repeatedly and independently until it comes up heads for the first time, then the expected number of coin flips is equal to 2.

6.7 The Binomial Distribution

As in Section 6.6, we consider an experiment that is successful with probability p and fails with probability $1 - p$. For an integer $n \geq 1$, we repeat the experiment, independently, n times. What is the expected number of times the experiment is successful?

We again model this problem using a coin that comes up heads with probability p and, thus, comes up tails with probability $1-p$. We flip the coin, independently, n times and define the random variable X to be the number of times the coin comes up heads. We want to determine the expected value $\mathbb{E}(X)$ of X.

Since our coin comes up heads with probability p, it is reasonable to guess that $\mathbb{E}(X)$ is equal to pn. For example, if $p = 1/2$, then, on average, half of the coin flips should come up heads. We will prove below that $\mathbb{E}(X)$ is indeed equal to pn.

6.7.1 Determining the Expected Value

Since the random variable X can take any value in $\{0, 1, 2, \ldots, n\}$, we have, by Lemma 6.4.2,

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot \Pr(X = k).$$

Thus, we have to determine $\Pr(X = k)$, i.e., the probability that in a sequence of n independent coin flips, the coin comes up heads exactly k times.

To give an example, assume that $n = 4$ and $k = 2$. The table below gives all $\binom{4}{2} = 6$ sequences of 4 coin flips that contain exactly 2 H's, together with their probabilities:

sequence	probability
$HHTT$	$p \cdot p \cdot (1-p) \cdot (1-p) = p^2(1-p)^2$
$HTHT$	$p \cdot (1-p) \cdot p \cdot (1-p) = p^2(1-p)^2$
$HTTH$	$p \cdot (1-p) \cdot (1-p) \cdot p = p^2(1-p)^2$
$THHT$	$(1-p) \cdot p \cdot p \cdot (1-p) = p^2(1-p)^2$
$THTH$	$(1-p) \cdot p \cdot (1-p) \cdot p = p^2(1-p)^2$
$TTHH$	$(1-p) \cdot (1-p) \cdot p \cdot p = p^2(1-p)^2$

As can be seen from this table, each of the $\binom{4}{2}$ sequences has the same probability $p^2(1-p)^2$. It follows that, if $n = 4$,

$$\Pr(X = 2) = \binom{4}{2} p^2(1-p)^2.$$

We now consider the general case. Let $n \geq 1$ and k be integers with $0 \leq k \leq n$. Then, $X = k$ if and only if there are exactly k H's in the sequence of n coin flips. The number of such sequences is equal to $\binom{n}{k}$, and each one of them has probability $p^k(1-p)^{n-k}$. Therefore, we have

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{6.3}$$

As a sanity check, let us use Newton's Binomial Theorem (i.e., Theorem 3.6.5) to verify that all probabilities add up to 1:

$$\begin{aligned} \sum_{k=0}^{n} \Pr(X = k) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\ &= ((1-p) + p)^n \\ &= 1. \end{aligned}$$

We are now ready to compute the expected value of the random variable X:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} k \cdot \Pr(X = k) \\ &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.\end{aligned}$$

Since

$$\begin{aligned}k\binom{n}{k} &= k \cdot \frac{n!}{k!(n-k)!} \\ &= n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} \\ &= n\binom{n-1}{k-1},\end{aligned}$$

we get

$$\mathbb{E}(X) = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

By changing the summation variable from k to $\ell + 1$, we get

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\ell=0}^{n-1} n \binom{n-1}{\ell} p^{\ell+1} (1-p)^{n-1-\ell} \\ &= pn \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{n-1-\ell}.\end{aligned}$$

By Newton's Binomial Theorem (i.e., Theorem 3.6.5), the summation is equal to

$$((1-p)+p)^{n-1} = 1.$$

Therefore, we get

$$\begin{aligned}\mathbb{E}(X) &= pn \cdot 1 \\ &= pn.\end{aligned}$$

We have done the following: Our intuition told us that $\mathbb{E}(X) = pn$. Then, we went through a painful calculation to show that our intuition was correct. There must be an easier way to show that $\mathbb{E}(X) = pn$. We will show below that there is indeed a much easier way.

6.7.2 Using the Linearity of Expectation

We define a sequence $X_1, X_2, \ldots, X_n$ of random variables as follows: For each i with $1 \leq i \leq n$,

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th coin flip results in heads,} \\ 0 & \text{if the } i\text{-th coin flip results in tails.} \end{cases}$$

Observe that

$$X = X_1 + X_2 + \cdots + X_n,$$

because

- X counts the number of H's in the sequence of n coin flips, and
- the summation on the right-hand side is equal to the number of 1's in the sequence $X_1, X_2, \ldots, X_n$, which, by definition, is equal to the number of H's in the sequence of n coin flips.

Using the Linearity of Expectation (see Theorem 6.5.2), we get

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \mathbb{E}\left(X_i\right). \end{aligned}$$

Thus, we have to determine the expected value of X_i. Since X_i is either 1 or 0, we have

$$\begin{aligned} \mathbb{E}\left(X_i\right) &= 1 \cdot \Pr\left(X_i = 1\right) + 0 \cdot \Pr\left(X_i = 0\right) \\ &= \Pr\left(X_i = 1\right) \\ &= \Pr(\text{the } i\text{-th coin flip results in heads}) \\ &= p. \end{aligned}$$

We conclude that

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{n} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n} p \\ &= pn.\end{aligned}$$

I hope you agree that this is much easier than what we did before.

The distribution function of the random variable X is given by (6.3). This function is called a binomial distribution:

Definition 6.7.1 Let $n \geq 1$ be an integer and let p be a real number with $0 < p < 1$. A random variable X has a *binomial distribution with parameters* n and p, if its distribution function satisfies

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for any integer k with $0 \leq k \leq n$.

Our calculation that led to the value of $\mathbb{E}(X)$ proves the following theorem:

Theorem 6.7.2 *Let $n \geq 1$ be an integer, let p be a real number with $0 < p < 1$, and let X be a random variable that has a binomial distribution with parameters n and p. Then*

$$\mathbb{E}(X) = pn.$$

6.8 Indicator Random Variables

In Section 6.7, we considered the random variable X whose value is equal to the number of heads in a sequence of n independent coin flips. In Section 6.7.2, we defined a sequence $X_1, X_2, \ldots, X_n$ of random variables, where $X_i = 1$ if the i-th coin flip results in heads and $X_i = 0$ otherwise. This random variable X_i *indicates* whether or not the i-th flip in the sequence is heads. Because of this, we call X_i an indicator random variable.

Definition 6.8.1 A random variable X is an *indicator random variable*, if it can only take values in $\{0, 1\}$.

As we have already seen in Section 6.7.2, the expected value of an indicator random variable is easy to determine:

Lemma 6.8.2 *If X is an indicator random variable, then*

$$\mathbb{E}(X) = \Pr(X = 1).$$

Proof. Since X is either 1 or 0, we have

$$\begin{aligned} \mathbb{E}(X) &= 1 \cdot \Pr(X = 1) + 0 \cdot \Pr(X = 0) \\ &= \Pr(X = 1). \end{aligned}$$

■

In the following subsections, we will see some examples of how indicator random variables can be used to compute the expected value of non-trivial random variables.

6.8.1 Runs in Random Bitstrings

Let n be a large integer. We generate a random bitstring

$$S = s_1 s_2 \ldots s_n$$

by flipping a fair coin, independently, n times. Let $k \geq 1$ be an integer. Recall from Section 5.13 that a *run of length k* is a consecutive subsequence of S, all of whose bits are equal. Define the random variable X to be the number of runs of length k.

For example, if S is the bitstring

0	0	1	1	1	1	1	0	0	0	1	1	0	0	0	0
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

and $k = 3$, then $X = 6$, because S contains 6 runs of length 3, starting at positions 3, 4, 5, 8, 13, and 14.

We want to determine the expected value $\mathbb{E}(X)$ of X.

A run of length k can start at any of the positions $1, 2, \ldots, n-k+1$. Our approach will be to define an indicator random variable that tells us whether or not the subsequence of length k that starts at any such position is a run. Thus, for any i with $1 \leq i \leq n-k+1$, we define the indicator random variable

$$X_i = \begin{cases} 1 & \text{if the subsequence } s_i s_{i+1} \ldots s_{i+k-1} \text{ is a run,} \\ 0 & \text{otherwise.} \end{cases}$$

Using Lemma 6.8.2, we get

$$\mathbb{E}(X_i) = \Pr(X_i = 1).$$

Since $X_i = 1$ if and only if all bits in the subsequence $s_i s_{i+1} \ldots s_{i+k-1}$ are 0 or all bits in this subsequence are 1, we have

$$\begin{aligned} \mathbb{E}(X_i) &= \Pr(X_i = 1) \\ &= (1/2)^k + (1/2)^k \\ &= 1/2^{k-1}. \end{aligned}$$

Since

$$X = \sum_{i=1}^{n-k+1} X_i,$$

the Linearity of Expectation (see Theorem 6.5.2) implies that

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n-k+1} X_i\right) \\ &= \sum_{i=1}^{n-k+1} \mathbb{E}(X_i) \\ &= \sum_{i=1}^{n-k+1} 1/2^{k-1} \\ &= \frac{n-k+1}{2^{k-1}}. \end{aligned}$$

Observe that the random variables $X_1, X_2, \ldots, X_{n+k-1}$ are not independent. (Can you see why?) Nevertheless, our derivation is correct, because the Linearity of Expectation is valid for any sequence of random variables.

For example, if we take $k = 1 + \log n$, then $2^{k-1} = 2^{\log n} = n$, so that

$$\mathbb{E}(X) = \frac{n - \log n}{n} = 1 - \frac{\log n}{n}.$$

Thus, for large values of n, the expected number of runs of length $1 + \log n$ is very close to 1. This is in line with Section 5.13, because we proved there that it is very likely that the sequence contains a run of length about $\log n$.

If we take $k = 1 + \frac{1}{2}\log n$, then

$$2^{k-1} = 2^{(\log n)/2} = 2^{\log \sqrt{n}} = \sqrt{n}$$

and

$$\mathbb{E}(X) = \frac{n - \frac{1}{2}\log n}{\sqrt{n}} = \sqrt{n} - \frac{\log n}{2\sqrt{n}}.$$

Thus, for large values of n, the expected number of runs of length $1 + \frac{1}{2}\log n$ is very close to $\sqrt{n}$.

6.8.2 Largest Elements in Prefixes of Random Permutations

Consider a sequence $s_1, s_2, \ldots, s_n$ of n numbers. The following algorithm computes the largest element in this sequence:

Algorithm FINDMAX$(s_1, s_2, \ldots, s_n)$**:**

```
max = −∞;
for i = 1 to n
do if s_i > max
    then max = s_i        (*)
    endif
endfor;
return max
```

We would like to know the number of times that line (*) is executed, i.e., the number of times that the value of the variable *max* changes. For example, if the input sequence is

$$3, 2, 5, 4, 6, 1,$$

then the value of *max* changes 3 times, namely when we encounter 3, 5, and 6. On the other hand, for the sequence

$$6, 5, 4, 3, 2, 1,$$

the value of *max* changes only once, whereas for

$$1, 2, 3, 4, 5, 6,$$

it changes 6 times.

Assume that the input sequence $s_1, s_2, \ldots, s_n$ is a random permutation of the integers $1, 2, \ldots, n$; thus, each permutation has probability $1/n!$ of being the input. We define the random variable X to be the number of times that line (*) is executed when running algorithm FindMax$(s_1, s_2, \ldots, s_n)$. We are interested in the expected value $\mathbb{E}(X)$ of X.

The algorithm makes n iterations and in each one, line (*) is either executed or not executed. We define, for each iteration, an indicator random variable that tells us whether or not line (*) is executed during that iteration. That is, for any i with $1 \leq i \leq n$, we define

$$X_i = \begin{cases} 1 & \text{if line (*) is executed in the } i\text{-th iteration,} \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$X = \sum_{i=1}^{n} X_i,$$

it follows from the Linearity of Expectation (see Theorem 6.5.2) that

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \\ &= \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) \\ &= \sum_{i=1}^{n} \Pr\left(X_i = 1\right). \end{aligned}$$

How do we determine $\Pr\left(X_i = 1\right)$? Observe that $X_i = 1$ if and only if the maximum of the subsequence $s_1, s_2, \ldots, s_i$ is at the last position in this

subsequence. Since the entire sequence $s_1, s_2, \ldots, s_n$ is a random permutation of $1, 2, \ldots, n$, the elements in the subsequence $s_1, s_2, \ldots, s_i$ are in a random order as well. The largest element in this subsequence is in any of the i positions with equal probability $1/i$. In particular, the probability that the largest element is at the last position in this subsequence is equal to $1/i$. It follows that

$$\Pr(X_i = 1) = 1/i.$$

Thus,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^{n} \Pr(X_i = 1) \\ &= \sum_{i=1}^{n} 1/i \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}. \end{aligned}$$

The number on the right-hand side is called the *harmonic number* and denoted by H_n. In the following subsection, we will show that H_n is approximately equal to $\ln n$. Thus, the expected number of times that line (*) of algorithm FindMax is executed, when given as input a random permutation of $1, 2, \ldots, n$, is about $\ln n$.

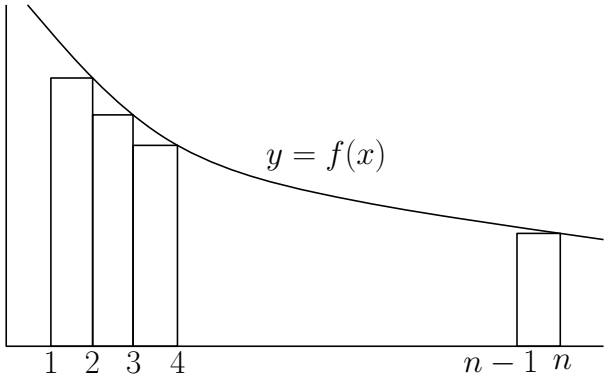
6.8.3 Estimating the Harmonic Number

Consider a positive real-valued decreasing function f. Thus, if $x \leq x'$, then $f(x) \geq f(x')$. We would like to estimate the summation

$$\sum_{i=1}^{n} f(i).$$

For example, if we take $f(x) = 1/x$, then the summation is the harmonic number H_n of the previous subsection.

For each i with $2 \leq i \leq n$, draw the rectangle with bottom-left corner at the point $(i-1, 0)$ and top-right corner at the point $(i, f(i))$, as in the figure below.



The area of the i-th rectangle is equal to $f(i)$ and, thus,

$$\sum_{i=1}^{n} f(i)$$

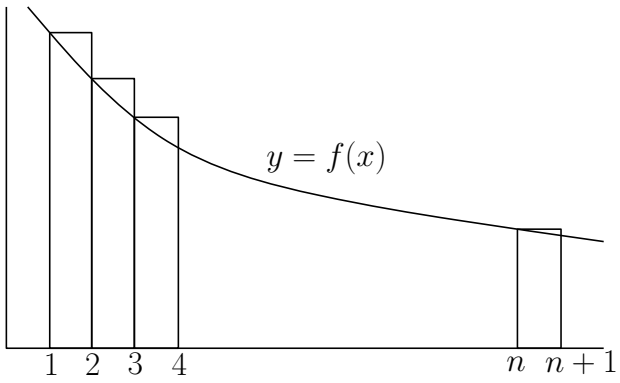
is equal to the sum of

- $f(1)$ and
- the total area of the $n-1$ rectangles.

Since f is decreasing, the rectangles are below the graph $y = f(x)$. It follows that the total area of the $n-1$ rectangles is less than or equal to the area between f and the x-axis, between $x = 1$ and $x = n$. We conclude that

$$\sum_{i=1}^{n} f(i) \leq f(1) + \int_{1}^{n} f(x)\,dx. \tag{6.4}$$

To obtain a lower bound on the summation, we modify the figure as indicated below: For each i with $1 \leq i \leq n$, we draw the rectangle with bottom-left corner at the point $(i, 0)$ and top-right corner at the point $(i + 1, f(i))$.



In this case, the graph $y = f(x)$ is below the top sides of the rectangles and, therefore,

$$\sum_{i=1}^{n} f(i) \geq \int_{1}^{n+1} f(x)\,dx. \tag{6.5}$$

If we apply (6.4) and (6.5) to the function $f(x) = 1/x$, then we get

$$\begin{aligned} H_n &= \sum_{i=1}^{n} \frac{1}{i} \\ &\leq 1 + \int_{1}^{n} \frac{dx}{x} \\ &= 1 + \ln n \end{aligned}$$

and

$$\begin{aligned} H_n &= \sum_{i=1}^{n} \frac{1}{i} \\ &\geq \int_{1}^{n+1} \frac{dx}{x} \\ &= \ln(n+1). \end{aligned}$$

We have proved the following result:

Lemma 6.8.3 *The harmonic number $H_n = \sum_{i=1}^{n} 1/i$ satisfies*

$$\ln(n+1) \leq H_n \leq 1 + \ln n.$$

6.9 The Insertion-Sort Algorithm

InsertionSort is a simple sorting algorithm that takes as input an array $A[1 \ldots n]$ of numbers. The algorithm uses a for-loop in which a variable i runs from 2 to n. At the start of the i-th iteration,

- the subarray $A[1 \ldots i-1]$ is sorted, whereas
- the algorithm has not yet seen any of the elements in the subarray $A[i \ldots n]$.

In the i-th iteration, the algorithm takes the element $A[i]$ and repeatedly swaps it with its left neighbor until the subarray $A[1 \ldots i]$ is sorted. The pseudocode of this algorithm is given below.

Algorithm InsertionSort($A[1 \ldots n]$)**:**

```
for i = 2 to n
do j = i;
    while j > 1 and A[j] < A[j − 1]
    do swap A[j] and A[j − 1];
        j = j − 1
    endwhile
endfor
```

We are interested in the total number of swaps that are made by this algorithm. The worst-case happens when the input array is sorted in reverse order, in which case the total number of swaps is equal to

$$1 + 2 + 3 + \cdots + (n-1) = \binom{n}{2}.$$

Thus, in the worst case, each of the $\binom{n}{2}$ pairs of input elements is swapped.

Assume that the input array $A[1 \ldots n]$ contains a uniformly random permutation of the integers $1, 2, \ldots, n$. Thus, each permutation has probability $1/n!$ of being the input. We define the random variable X to be the total number of swaps made when running algorithm InsertionSort($A[1 \ldots n]$). We will determine the expected value $\mathbb{E}(X)$ of X.

Since we want to count the number of pairs of input elements that are swapped, we will use, for each pair of input elements, an indicator random

variable that indicates whether or not this pair gets swapped by the algorithm. That is, for each a and b with $1 \leq a < b \leq n$, we define

$$X_{ab} = \begin{cases} 1 & \text{if } a \text{ and } b \text{ get swapped by the algorithm,} \\ 0 & \text{otherwise.} \end{cases}$$

We observe that, since $a < b$, these two elements get swapped if and only if in the input array, b is to the left of a. Since the input array contains a uniformly random permutation, the events "b is to the left of a" and "a is to the left of b" are symmetric. Therefore, we have

$$\mathbb{E}\left(X_{ab}\right) = \Pr\left(X_{ab} = 1\right) = 1/2.$$

This can also be proved by showing that there are $n!/2$ permutations of $1, 2, \ldots, n$ in which b appears to the left of a and, thus, $n!/2$ permutations in which a appears to the left of b.

Since each pair of input elements is swapped at most once, we have

$$X = \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}.$$

It follows from the Linearity of Expectation (see Theorem 6.5.2) that

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}\right) \\ &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \mathbb{E}\left(X_{ab}\right) \\ &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \frac{1}{2} \\ &= \frac{1}{2}\binom{n}{2}. \end{aligned}$$

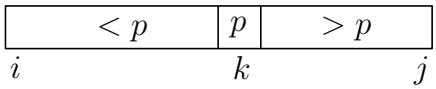
Thus, the expected number of swaps on a random input array is one half times the worst-case number of swaps.

6.10 The Quick-Sort Algorithm

We have already seen algorithm QuickSort in Section 1.3. This algorithm takes as input an array $A[1 \ldots n]$ of numbers, which we assume for simplicity to be distinct. A generic call QuickSort(A, i, j) takes two indices i and j and sorts the subarray $A[i \ldots j]$. Thus, the call QuickSort$(A, 1, n)$ sorts the entire array.

Algorithm QuickSort$(A, i, j]$**):**

if $i < j$
then p = uniformly random element in $A[i \ldots j]$;
compare p with all other elements in $A[i \ldots j]$;
rearrange $A[i \ldots j]$ such that it has the following form (this rearranging defines the value of k):



QuickSort$(A, i, k-1)$;
QuickSort$(A, k+1, j)$
endif

The element p is called the *pivot*. We have seen in Section 1.3 that the worst-case running time of algorithm QuickSort$(A, 1, n)$ is $\Theta(n^2)$. In this section, we will prove that the expected running time is only $O(n \log n)$.

We assume for simplicity that the input array is a permutation of the integers $1, 2, \ldots, n$. We do not make any other assumption about the input. In particular, we do not assume that the input is a random permutation. The only place where randomization is used is when the pivot is chosen: It is chosen uniformly at random in the subarray on which QuickSort is called.

We define the random variable X to be the total number of comparisons between pairs of input elements that are made by algorithm QuickSort$(A, 1, n)$. We will prove that $\mathbb{E}(X) = O(n \log n)$.

As in Section 6.9, we define for each a and b with $1 \leq a < b \leq n$, the

indicator random variable

$$X_{ab} = \begin{cases} 1 & \text{if } a \text{ and } b \text{ are compared to each other when} \\ & \text{running } \textsc{QuickSort}(A, 1, n), \\ 0 & \text{otherwise.} \end{cases}$$

Since each pair of input elements is compared at most once, we have

$$X = \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}.$$

It follows from the Linearity of Expectation (see Theorem 6.5.2) that

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{a=1}^{n-1} \sum_{b=a+1}^{n} X_{ab}\right) \\ &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \mathbb{E}\left(X_{ab}\right) \\ &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \Pr\left(X_{ab} = 1\right). \end{aligned}$$

We consider two input elements a and b with $1 \leq a < b \leq n$. We are going to determine $\Pr\left(X_{ab} = 1\right)$, i.e., the probability that the elements a and b are compared to each other when running algorithm $\textsc{QuickSort}(A, 1, n)$. Define the set

$$S_{ab} = \{a, a+1, \ldots, b\}.$$

At the start of algorithm $\textsc{QuickSort}(A, 1, n)$, all elements of the set S_{ab} are part of the input. Consider the first pivot p that is chosen. We observe the following:

- Assume that $p \notin S_{ab}$.
  - If $p < a$, then after the algorithm has rearranged the input array, all elements of the set S_{ab} are to the right of p and, thus, all these elements are part of the input for the recursive call $\textsc{QuickSort}(A, k+1, n)$. During the rearranging, a and b were not compared to each other. However, they *may be* compared during later recursive calls.

  - If $p > b$, then after the algorithm has rearranged the input array, all elements of the set S_{ab} are to the left of p and, thus, all these elements are part of the input for the recursive call QuickSort$(A, 1, k-1)$. During the rearranging, a and b were not compared to each other. However, they *may be* compared during later recursive calls.

- Assume that $p \in S_{ab}$.

  - If $p \neq a$ and $p \neq b$, then after the algorithm has rearranged the input array, a is to the left of p and b is to the right of p. During the rearranging, a and b were not compared to each other. Also, since a and b have been "separated", they will not be compared during later recursive calls. Thus, we have $X_{ab} = 0$.
  - If $p = a$ or $p = b$, then during the rearranging, a and b are compared to each other. Thus, we have $X_{ab} = 1$. (Note that in later recursive calls, a and b will not be compared to each other again.)

We conclude that whether or not a and b are compared to each other is completely determined by the element p of the set S_{ab} that is the first element in this set to be chosen as a pivot. If this element p is equal to a or b, then $X_{ab} = 1$. On the other hand, if this element p belongs to $S_{ab} \setminus \{a, b\}$, then $X_{ab} = 0$. Since

- in any recursive call, the pivot is chosen uniformly at random from the subarray that is the input for this call and
- at the start of the first recursive call in which the pivot belongs to the set S_{ab}, all elements of this set are part of the input for this call,

each of the $b - a + 1$ elements of S_{ab} has the same probability of being the first element of S_{ab} that is chosen as a pivot. It follows that

$$\Pr(X_{ab} = 1) = \frac{2}{b - a + 1}.$$

We conclude that

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \Pr\left(X_{ab} = 1\right) \\
&= \sum_{a=1}^{n-1} \sum_{b=a+1}^{n} \frac{2}{b-a+1} \\
&= 2 \sum_{a=1}^{n-1} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-a+1} \right) \\
&\leq 2 \sum_{a=1}^{n-1} \left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \right) \\
&= 2 \sum_{a=1}^{n-1} \left(H_n - 1 \right) \\
&= 2(n-1) \left(H_n - 1 \right) \\
&\leq 2n \left(H_n - 1 \right),
\end{aligned}$$

where H_n is the harmonic number. Using Lemma 6.8.3, it follows that

$$\mathbb{E}(X) \leq 2n \ln n.$$

6.11 Skip Lists

Consider a set S of n numbers. We would like to store these numbers in a data structure that supports the following operations:

- SEARCH(x): This operation returns the largest element in the set S that is less than or equal to x.
- INSERT(x): This operation inserts the number x into the set S.
- DELETE(x): This operation deletes the number x from the set S.

A standard data structure for this problem is a balanced binary search tree (such as a red-black tree or an AVL-tree), which allows each of these three operations to be performed in $O(\log n)$ time. Searching in a binary search tree is straightforward, but keeping the tree balanced after an insertion or deletion is cumbersome.

In this section, we introduce *skip lists* as an alternative data structure. A skip list is constructed using the outcomes of coin flips, which result in a structure that is balanced in the expected sense. Because of this, the insertion and deletion algorithms become straightforward: We, as a programmer, do not have to take care of rebalancing operations, because the coin flips take care of this.

To define a skip list for the set S, we first construct a sequence $S_0, S_1, S_2, \ldots$ of subsets of S:

- Let $S_0 = S$.
- For $i = 0, 1, 2, \ldots$, assume that the set S_i has already been constructed. If S_i is non-empty, we do the following:
  - Initialize an empty set S_{i+1}.
  - For each element y in the set S_i, flip a fair and independent coin. If the coin comes up heads, element y is added to the set S_{i+1}.

The process terminates as soon as the next set S_{i+1} is empty.

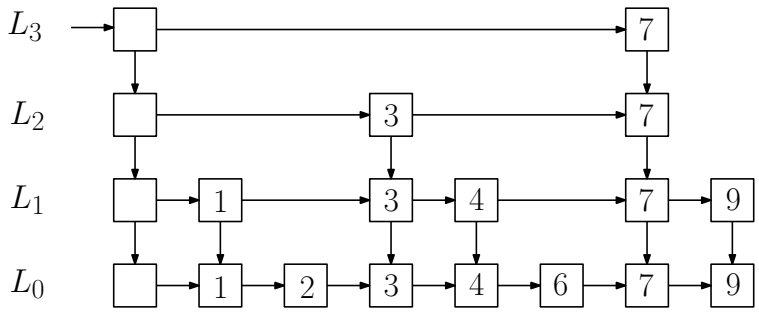
Let h be the number of non-empty sets that are constructed by this process, and consider the sequence $S_0, S_1, \ldots, S_h$ of sets. Observe that h is a random variable and each of the sets $S_1, S_2, \ldots, S_h$ is a random subset of S.

The skip list for S consists of the following:

- For each i with $0 \leq i \leq h$, we store the sorted sequence of elements of the set S_i in a linked list L_i.
  - Each node u of L_i stores one element of S_i, which is denoted by $key(u)$.
  - Each node u of L_i stores a pointer to its successor node in L_i, which is denoted by $right(u)$. If u is the rightmost node in L_i, then $right(u) = nil$.
  - We add a *dummy node* at the beginning of L_i. The key of this node is *nil* and its successor is the node of L_i whose key is the smallest element in S_i.
- For each i with $1 \leq i \leq h$ and each node u of L_i, u stores a pointer to the node u' in L_{i-1} for which $key(u') = key(u)$. The node u' is denoted by $down(u)$.

- There is a pointer to the dummy node in the list L_h. We will refer to this node as the *root* of the skip list.

The value of h is called the *height* of the skip list. An example of a skip list of height 3 for the set $S = \{1, 2, 3, 4, 6, 7, 9\}$ is shown in the figure below.



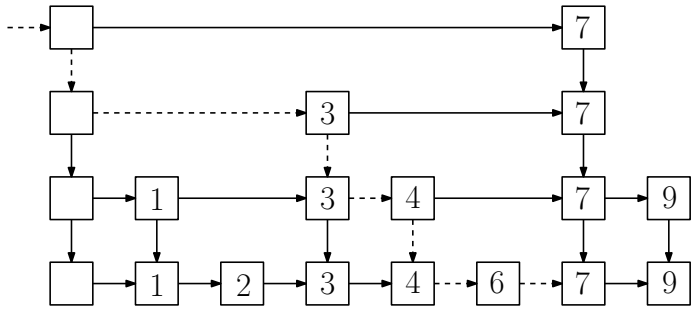
6.11.1 Algorithm SEARCH

The algorithm that searches for a number x keeps track of the current node u and the index i of the list L_i that contains u. Initially, u is the root of the skip list and $i = h$. At any moment, if $i \geq 1$, the algorithm tests if the key of $right(u)$ is less than x. If this is the case, then u moves one node to the right in the list L_i; otherwise, u moves to the node $down(u)$ in the list L_{i-1}. Once $i = 0$, node u moves to the right in the list L_0 and stops at the last node whose key is at most equal to x. The pseudocode of this algorithm SEARCH(x) is given below.

Algorithm SEARCH(x)**:**

// returns the rightmost node u in L_0 such that $key(u) \leq x$
u = root of the skip list;
$i = h$;
while $i \geq 1$
do if $right(u) \neq nil$ and $key(right(u)) < x$
  then $u = right(u)$
  else $u = down(u)$;
    $i = i - 1$
  endif
endwhile;
while $right(u) \neq nil$ and $key(right(u)) \leq x$
do $u = right(u)$
endwhile;
return u

The dashed edges in the figure below show the path that is followed when running algorithm SEARCH(7). Note that if we replace "$key(right(u)) < x$" in the first while-loop by "$key(right(u)) \leq x$", we obtain a different path that ends in the same node: This path moves from the root to the node in L_3 whose key is 7, and then it moves down to the list L_0. As we will see later, using the condition "$key(right(u)) < x$" simplifies the algorithm for deleting an element from the skip list.



6.11.2 Algorithms INSERT and DELETE

Algorithm INSERT(x) takes as input a number x and inserts it into the skip list. This algorithm works as follows:

- Run algorithm SEARCH(x) and consider the node u that is returned. We assume that $key(u) \neq x$ and, thus, x is not in the skip list yet. Observe that the new element x belongs between the nodes u and $right(u)$.

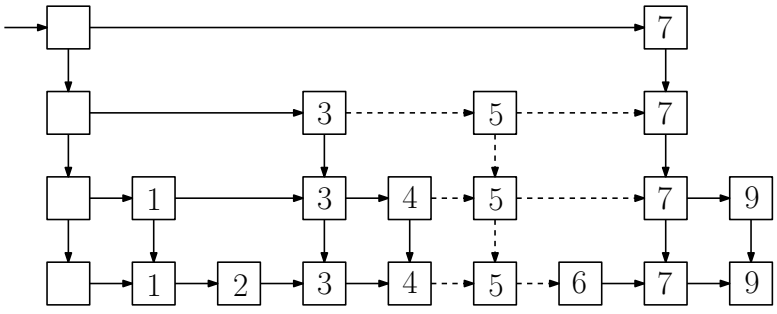
- Flip a fair and independent coin repeatedly until it comes up tails for the first time. Let k be the number of flips.

- Add the new element x to the lists $L_0, L_1, \ldots, L_{k-1}$. Note that if $k \geq h+2$, we have to add new lists $L_{h+1}, \ldots, L_{k-1}$ to the skip list (each one containing a dummy node and a node storing x), set $h = k - 1$, and update the pointer to the root of the new skip list.

- When adding x to a list L_i, we have to know its predecessor in this list.

  - To find these predecessors, we modify algorithm SEARCH(x) as follows: Each time the current node u moves down, we push u onto an initially empty stack. In this way, the predecessors that we need are stored, in the correct order, on the stack.
  - An easier way that avoids using a stack is to flip the coin and, thus, determine k, before running algorithm SEARCH(x). We then modify algorithm SEARCH(x): If $i < k$ and the current node u moves down, we add the new element x to L_i between the nodes u and $right(u)$.

The figure below shows the skip list that results when inserting the number 5 into the skip list of the previous figure. In this case, $k = 3$ and the new element is added to the lists L_0, L_1, and L_2. The dashed edges show the pointers that are changed during this insertion.



Algorithm DELETE(x) takes as input a number x and deletes it from the skip list. This algorithm does the following:

- Run a modified version of algorithm SEARCH(x): Each time the current node u moves down, test if $key(right(u)) = x$. If this is the case, delete the node $right(u)$ by setting $right(u) = right(right(u))$. Finally, delete the node in L_0 whose key is equal to x.
- At this moment, it may happen that some of the lists $L_h, L_{h-1}, \ldots$ only consist of dummy nodes. If this is the case, delete these lists, and update the height h and the root of the new skip list.

Implementation details of skip lists and algorithms SEARCH, INSERT, and DELETE can be found in Pat Morin's free textbook *Open Data Structures*, which is available at `http://opendatastructures.org/`

6.11.3 Analysis of Skip Lists

In this subsection, we will prove that the expected size of a skip list is $O(n)$ and the expected running time of algorithm SEARCH is $O(\log n)$. This will imply that the expected running times of algorithms INSERT and DELETE are $O(\log n)$ as well. Throughout this subsection, we assume for simplicity that n is a power of 2, so that $\log n$ is an integer.

Consider again the lists $L_0, L_1, \ldots, L_h$ in the skip list. For the purpose of analysis, we define for each integer $i > h$, L_i to be an empty list.

For each element x in the list L_0, we define the random variable $h(x)$ to be the largest value of i such that x is contained in the list L_i. Thus, x occurs in the lists $L_0, L_1, \ldots, L_{h(x)}$, but not in the list $L_{h(x)+1}$.

Lemma 6.11.1 *For any element x in the list L_0, we have*

$$\mathbb{E}(h(x)) = 1.$$

Proof. The value of $h(x)$ is determined by the following process: flip a fair coin repeatedly and independently until it comes up tails for the first time. The value of $h(x)$ is then equal to the number of flips minus one. For example, if we flip the coin three times (i.e., obtain the sequence HHT), then x is contained in the lists L_0, L_1, and L_2, but not in L_3; thus, $h(x) = 2$. By Theorem 6.6.2, the expected number of coin flips is equal to two. As a result, the expected value of $h(x)$ is equal to one. ∎

Lemma 6.11.2 *For any element x in the list L_0 and for any $i \geq 0$,*

$$\Pr(x \in L_i) = 1/2^i.$$

Proof. The claim follows from the fact that x is contained in the list L_i if and only if the first i coin flips for x all result in heads. ∎

Lemma 6.11.3 *Ignoring the dummy element, we have, for each $i \geq 0$,*

$$\mathbb{E}(|L_i|) = n/2^i.$$

Proof. We know from Lemma 6.11.2 that each element x is contained in L_i with probability $1/2^i$, independently of the other elements in L_i. Therefore, the size of L_i is a random variable that has a binomial distribution with parameters n and $p = 1/2^i$. The claim then follows from Theorem 6.7.2. ∎

Lemma 6.11.4 *Let X be the random variable that is equal to the total number of nodes in all lists $L_0, L_1, L_2, \ldots$, ignoring the dummy nodes. Then*

$$\mathbb{E}(X) = 2n.$$

Proof. We will give two proofs. In the first proof, we use the fact that

$$X = \sum_{i=0}^{\infty} |L_i|.$$

Using the Linearity of Expectation (i.e., Theorem 6.5.3) and Lemmas 6.11.3 and 5.14.2, we get

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=0}^{\infty} |L_i|\right) \\
&= \sum_{i=0}^{\infty} \mathbb{E}(|L_i|) \\
&= \sum_{i=0}^{\infty} n/2^i \\
&= n \sum_{i=0}^{\infty} (1/2)^i \\
&= 2n.
\end{aligned}$$

In the second proof, we use the fact that each element x occurs in exactly $h(x)+1$ lists, namely $L_0, L_1, \ldots, L_{h(x)}$. Thus, we have

$$X = \sum_x (h(x)+1).$$

Using the Linearity of Expectation (i.e., Theorem 6.5.2) and Lemma 6.11.1, we get

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\sum_x (h(x)+1)\right) \\
&= \sum_x \mathbb{E}(h(x)+1) \\
&= \sum_x (\mathbb{E}(h(x))+1) \\
&= \sum_x 2 \\
&= 2n.
\end{aligned}$$

■

Lemma 6.11.5 *The expected height h of the skip list satisfies*

$$\mathbb{E}(h) \leq \log n + 1.$$

Proof. Since

$$h = \max_x h(x),$$

we have

$$\mathbb{E}(h) = \mathbb{E}\left(\max_x h(x)\right).$$

It is tempting, but wrong, to think that this is equal to

$$\max_x \mathbb{E}(h(x)),$$

which is equal to 1 by Lemma 6.11.1. (In Exercise 6.27, you will find a simple example showing that, in general, the expected value of a maximum is not equal to the maximum of the expected values.)

To prove a correct upper bound on $\mathbb{E}(h)$, we introduce for each integer $i \geq 1$, an indicator random variable

$$X_i = \begin{cases} 1 & \text{if the list } L_i \text{ is non-empty,} \\ 0 & \text{otherwise.} \end{cases}$$

We observe that

$$h = \sum_{i=1}^{\infty} X_i.$$

Since X_i is always less than or equal to 1, it is obvious that

$$\mathbb{E}\left(X_i\right) \leq 1. \tag{6.6}$$

Also, since X_i is always less than or equal to the size $|L_i|$ of the list L_i (ignoring the dummy node), i.e., $X_i \leq |L_i|$, we have, using Lemma 6.11.3,

$$\mathbb{E}\left(X_i\right) \leq \mathbb{E}\left(|L_i|\right) = n/2^i. \tag{6.7}$$

Using the Linearity of Expectation (i.e., Theorem 6.5.3), we get

$$\begin{aligned} \mathbb{E}(h) &= \mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right) \\ &= \sum_{i=1}^{\infty} \mathbb{E}\left(X_i\right) \\ &= \sum_{i=1}^{\log n} \mathbb{E}\left(X_i\right) + \sum_{i=\log n+1}^{\infty} \mathbb{E}\left(X_i\right). \end{aligned}$$

If we apply (6.6) to the first summation and (6.7) to the second summation,

we get

$$\begin{aligned}
\mathbb{E}(h) &\leq \sum_{i=1}^{\log n} 1 + \sum_{i=\log n+1}^{\infty} \frac{n}{2^i} \\
&= \log n + \sum_{j=0}^{\infty} \frac{n}{2^{\log n+1+j}} \\
&= \log n + \sum_{j=0}^{\infty} \frac{n}{n \cdot 2^{1+j}} \\
&= \log n + \sum_{j=0}^{\infty} \frac{1}{2^{1+j}} \\
&= \log n + \frac{1}{2} \sum_{j=0}^{\infty} \frac{1}{2^j} \\
&= \log n + \frac{1}{2} \cdot 2 \\
&= \log n + 1.
\end{aligned}$$

■

Lemma 6.11.6 *Let Y be the random variable that is equal to the total number of nodes in all lists $L_0, L_1, L_2, \ldots$, including the dummy nodes. Then*

$$\mathbb{E}(Y) \leq 2n + \log n + 2.$$

Proof. The total number of dummy nodes is equal to $h + 1$. Using the notation of Lemma 6.11.4, we have

$$Y = X + h + 1.$$

Thus, using the Linearity of Expectation (i.e., Theorem 6.5.2) and Lemmas 6.11.4 and 6.11.5, we get

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(X + h + 1) \\
&= \mathbb{E}(X) + \mathbb{E}(h) + 1 \\
&\leq 2n + (\log n + 1) + 1 \\
&= 2n + \log n + 2.
\end{aligned}$$

■

Consider any number x. As we have seen in Section 6.11.1, algorithm SEARCH(x) starts at the root of the skip list and follows a path to the rightmost node u in the bottom list L_0 for which $key(u) \leq x$. We will refer to this path as the *search path* of the algorithm.

Lemma 6.11.7 *For any number x, let N be the random variable that is equal to the number of nodes on the search path of algorithm* SEARCH(x). *Then*

$$\mathbb{E}(N) \leq 2 \log n + 5.$$

Proof. Consider the node u that is returned by algorithm SEARCH(x), let v be the second last node on the search path, let P be the search path from the root to v, and let M be the random variable that is equal to the number of nodes on P. Then, $N = M + 1$ and

$$\mathbb{E}(N) = \mathbb{E}(M + 1) = \mathbb{E}(M) + 1.$$

It suffices to prove that

$$\mathbb{E}(M) \leq 2 \log n + 4.$$

Define the following path P':

- P' starts at node v.
- At any node w on P', the path P' goes up one level if possible, and goes left one node otherwise.

You should convince yourself that this path P' is the reverse of P and, therefore, M is the number of nodes on P'.

For each $i \geq 0$, define the random variable M_i to be equal to the number of nodes in the list L_i at which the path P' goes left. Then, M is the sum of

- h: these are the nodes on P' at which P' goes up one level,
- 1: this accounts for the last node on P', which is the root, and
- $\sum_{i=0}^{h} M_i$.

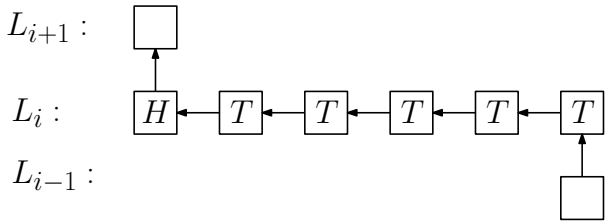
Thus,

$$\begin{aligned}\mathbb{E}(M) &= \mathbb{E}\left(h+1+\sum_{i=0}^{h} M_i\right) \\ &= \mathbb{E}(h)+1+\left(\sum_{i=0}^{h} M_i\right).\end{aligned}$$

Observe that the number of terms in the latter summation is equal to $h+1$, which is a random variable. In general, the Linearity of Expectation does not apply to summations of a random number of terms; see Exercise 6.28 for an example. Therefore, we proceed as follows. We first observe that

$$M = h+1+\sum_{i=0}^{\infty} M_i.$$

As the figure below indicates, the random variable M_i can be interpreted as being the number of tails obtained when flipping a fair coin until it comes up heads for the first time. Since (i) the list L_i may be empty (in which case $M_i = 0$) or (ii) the portion of the path P' in L_i may terminate because it reaches the dummy node, M_i is in fact less than or equal to the number of tails.



Therefore, by Theorem 6.6.2,

$$\mathbb{E}(M_i) \leq 1. \tag{6.8}$$

Also, since M_i is less than or equal to the size $|L_i|$ of the list L_i (ignoring the dummy node), we have, using Lemma 6.11.3,

$$\mathbb{E}(M_i) \leq \mathbb{E}(|L_i|) = n/2^i. \tag{6.9}$$

Using the Linearity of Expectation (i.e., Theorem 6.5.3), we get

$$\begin{aligned}\mathbb{E}(M) &= \mathbb{E}\left(h+1+\sum_{i=0}^{\infty} M_i\right)\\ &= \mathbb{E}(h)+1+\sum_{i=0}^{\infty}\mathbb{E}\left(M_i\right)\\ &= \mathbb{E}(h)+1+\sum_{i=0}^{\log n}\mathbb{E}\left(M_i\right)+\sum_{i=\log n+1}^{\infty}\mathbb{E}\left(M_i\right).\end{aligned}$$

We know from Lemma 6.11.5 that $\mathbb{E}(h) \leq \log n + 1$. If we apply (6.8) to the first summation and (6.9) to the second summation, we get

$$\begin{aligned}\mathbb{E}(M) &\leq (\log n+1)+1+\sum_{i=0}^{\log n} 1+\sum_{i=\log n+1}^{\infty} n/2^i\\ &= 2\log n+3+\sum_{i=\log n+1}^{\infty} n/2^i.\end{aligned}$$

We have seen the infinite series in the proof of Lemma 6.11.5 and showed that it is equal to 1. Thus, we conclude that

$$\mathbb{E}(M) \leq 2\log n + 4.$$

∎

6.12 Exercises

6.1 Consider a fair coin that has 0 on one side and 1 on the other side. We flip this coin once and roll a fair die twice. Define the following random variables:

$$\begin{aligned}X &= \text{the result of the coin},\\ Y &= \text{the sum of the two dice},\\ Z &= X \cdot Y.\end{aligned}$$

- Determine the distribution functions of X, Y, and Z.

- Are X and Y independent random variables?
- Are X and Z independent random variables?
- Are Y and Z independent random variables?
- Are X, Y and Z mutually independent random variables?

6.2 We flip a fair coin twice (independently). For each heads, you win 3 dollars, whereas for each tails, you lose 2 dollars. Define the random variable X to be the amount of money that you win.

- Use the definition of expected value to compute $\mathbb{E}(X)$.
- Use the Linearity of Expectation to compute $\mathbb{E}(X)$.

We flip a fair coin 27 times (independently). For each heads, you win 3 dollars, whereas for each tails, you lose 2 dollars. Define the random variable Y to be the amount of money that you win.

- Compute the expected value $\mathbb{E}(Y)$ of Y.

6.3 Let r and b be positive integers and define $\alpha = \frac{r}{r+b}$. A bowl contains r red balls and b blue balls; thus, α is the fraction of the balls that are red. Consider the following experiment:

- Choose one ball uniformly at random.
  - If the chosen ball is red, then put it back, together with an additional red ball.
  - If the chosen ball is blue, then put it back, together with an additional blue ball.

Define the random variable X to be the fraction of the balls that are red, after this experiment. Prove that $\mathbb{E}(X) = \alpha$.

6.4 The Ontario Lottery and Gaming Corporation (OLG) offers the following lottery game:

- OLG chooses a winning number w in the set $S = \{0, 1, 2, \ldots, 999\}$.
- If John wants to play, he pays \$1 and chooses a number x in S.

  - If $x = w$, then John receives \$700 from OLG. In this case, John wins \$699.
  - Otherwise, $x \neq w$ and John does not receive anything. In this case, John loses \$1.

Assume that

- John plays this game once per day for one year (i.e., for 365 days),
- each day, OLG chooses a new winning number,
- each day, John chooses x uniformly at random from the set S, independently from previous choices.

Define the random variable X to be the total amount of dollars that John wins during one year. Determine the expected value $\mathbb{E}(X)$. (*Hint:* Use the Linearity of Expectation.)

6.5 Assume we roll two fair and independent dice. Let (i, j) be the outcome, where i is the result of the first die and j is the result of the second die. Define the random variables

$$X = |i - j|$$

and

$$Y = \max(i, j).$$

Are X and Y independent random variables?

6.6 We flip a fair coin independently and stop as soon as we get a T followed by an H. Define the random variable X to be the number of coin flips. For example, if the coin flips are $HHHTTTTH$, then $X = 8$.

- Determine the expected value $\mathbb{E}(X)$ of X. (*Hint:* Use the Linearity of Expectation.)

6.7 In Section 6.6, we have shown that for $-1 < x < 1$,

$$\sum_{k=1}^{\infty} kx^k = \frac{x}{(1-x)^2}.$$

In this question, you will prove this formula in a different way.

Consider the following infinite matrix:

$$\begin{pmatrix} x & 0 & 0 & 0 & 0 & 0 & \dots \\ x^2 & x^2 & 0 & 0 & 0 & 0 & \dots \\ x^3 & x^3 & x^3 & 0 & 0 & 0 & \dots \\ x^4 & x^4 & x^4 & x^4 & 0 & 0 & \dots \\ x^5 & x^5 & x^5 & x^5 & x^5 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

We are going to add all elements in this matrix in two different ways. A *row-sum* is the sum of all elements in one row, whereas a *column-sum* is the sum of all elements in one column.

Note that the sum of all row-sums is equal to

$$x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \cdots = \sum_{k=1}^{\infty} kx^k.$$

- Using only the formula $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ and algebraic manipulation, prove that the sum of all column-sums is equal to

$$\frac{x}{(1-x)^2}.$$

6.8 Let X be a random variable that takes values in $\{0, 1, 2, 3, \dots\}$. By Lemma 6.4.2, we have

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \Pr(X = k).$$

Define an infinite matrix and use it to prove that

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \Pr(X \geq k).$$

6.9 Let $0 < p < 1$ and consider a coin that comes up heads with probability p and tails with probability $1 - p$. We flip the coin independently until it comes up heads for the first time. Define the random variable X to be the number of times that we flip the coin. In Section 6.6, we have shown that $\mathbb{E}(X) = 1/p$. Below, you will prove this in a different way.

- Let $k \geq 1$ be an integer. Determine $\Pr(X \geq k)$.
- Using only the formula $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$, the expression for $\mathbb{E}(X)$ from Exercise 6.8, and your answer for $\Pr(X \geq k)$, prove that $\mathbb{E}(X) = 1/p$.

6.10 The Ottawa Senators and the Toronto Maple Leafs play a best-of-seven series: These two hockey teams play against each other until one of them has won four games. Assume that

- in any game, the Sens have a probability of 3/4 of defeating the Leafs,
- the results of the games are independent.

Determine the probability that seven games are played in this series.

6.11 Consider an experiment that is successful with probability 0.8. We repeat this experiment (independently) until it is successful for the first time. The first 5 times we do the experiment, we have to pay \$10 per experiment. After this, we have to pay \$5 per experiment. Define the random variable X to be the total amount of money that we have to pay during all experiments. Determine the expected value $\mathbb{E}(X)$.

Hint: Recall that $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1-x)^2$.

6.12 When Lindsay and Simon have a child, this child is a boy with probability 1/2 and a girl with probability 1/2, independent of the gender of previous children. Lindsay and Simon stop having children as soon as they have a girl. Define the random variables

$$B = \text{the number of boys that Lindsay and Simon have}$$

and

$$G = \text{the number of girls that Lindsay and Simon have.}$$

Determine the expected values $\mathbb{E}(B)$ and $\mathbb{E}(G)$.

6.13 Let p be a real number with $0 < p < 1$. When Lindsay and Simon have a child, this child is a boy with probability p and a girl with probability $1 - p$, independent of the gender of previous children. Lindsay and Simon stop having children as soon as they have a child that has the same gender as their first child. Define the random variable X to be the number of children that Lindsay and Simon have. Determine the expected value $\mathbb{E}(X)$.

Hint: Recall that $\sum_{k=1}^{\infty} kx^{k-1} = 1/(1-x)^2$.

6.14 Let $X_1, X_2, \ldots, X_n$ be a sequence of mutually independent random variables. For each i with $1 \leq i \leq n$, assume that

- the variable X_i is either equal to 0 or equal to $n+1$, and
- $\mathbb{E}(X_i) = 1$.

Determine

$$\Pr(X_1 + X_2 + \cdots + X_n \leq n).$$

6.15 A *maximal run of ones* in a bitstring is a maximal consecutive substring of ones. For example, the bitstring 1000111110100111 has four maximal runs of ones: 1, 11111, 1, and 111.

Let $n \geq 1$ be an integer and consider a random bitstring of length n. Define the random variable X to be the number of maximal runs of ones in this bitstring. Determine the expected value $\mathbb{E}(X)$ of X. (*Hint:* Use indicator random variables.)

6.16 Let $A[1 \ldots n]$ be an array of n numbers. Consider the following two algorithms, which take as input the array A and a number x. If x is not present in A, then these algorithms return the message "not present". Otherwise, they return an index i such that $A[i] = x$. The first algorithm runs linear search from left to right, whereas the second algorithm runs linear search from right to left.

Algorithm LinearSearchLeftToRight(A, x)**:**

$i := 1$;
while $i \leq n$ and $A[i] \neq x$ **do** $i := i + 1$ **endwhile**;
if $i = n + 1$ **then** return "not present" **else** return i **endif**

Algorithm LinearSearchRightToLeft(A, x)**:**

$i := n$;
while $i \geq 1$ and $A[i] \neq x$ **do** $i := i - 1$ **endwhile**;
if $i = 0$ **then** return "not present" **else** return i **endif**

Consider the following algorithm, which again take as input the array A and a number x. If x is not present in A, then it returns the message "not present". Otherwise, it returns an index i such that $A[i] = x$.

Algorithm RandomLinearSearch(A, x)**:**

flip a fair coin;
if the coin comes up heads
then LinearSearchLeftToRight(A, x)
else LinearSearchRightToLeft(A, x)
endif

Assume that the number x occurs exactly once in the array A and let k be the index such that $A[k] = x$. Let X be the random variable whose value is the number of times the test "$A[i] \neq x$" is made in algorithm RandomLinearSearch(A, x). (In words, X is the number of comparisons made by algorithm RandomLinearSearch(A, x).) Determine the expected value $\mathbb{E}(X)$ of X.

6.17 Consider the set $V = \{1, 2, \ldots, n\}$ and let p be a real number with $0 < p < 1$. We construct a graph $G = (V, E)$ with vertex set V, whose edge set E is determined by the following random process: Each unordered pair $\{i, j\}$ of vertices, where $i \neq j$, occurs as an edge in E with probability p, independently of the other unordered pairs.

A *triangle* in G is an unordered triple $\{i, j, k\}$ of distinct vertices, such that $\{i, j\}$, $\{j, k\}$, and $\{k, i\}$ are edges in G.

Define the random variable X to be the total number of triangles in the graph G. Determine the expected value $\mathbb{E}(X)$. (*Hint:* Use indicator random variables.)

6.18 In Section 6.9, we have seen the following algorithm InsertionSort, which sorts any input array $A[1 \ldots n]$:

Algorithm InsertionSort$(A[1 \ldots n])$**:**

for $i = 2$ **to** n
do $j = i$;
  while $j > 1$ and $A[j] < A[j-1]$
  do swap $A[j]$ and $A[j-1]$;
    $j = j - 1$
  endwhile
endfor

Consider an input array $A[1 \ldots n]$, where each element $A[i]$ is chosen independently and uniformly at random from the set $\{1, 2, \ldots, m\}$.

- Let i and j be two indices with $1 \leq i < j \leq n$, and consider the values $A[i]$ and $A[j]$ (just before the algorithm starts). Prove that

$$\Pr(A[i] > A[j]) = \frac{1}{2} - \frac{1}{2m}.$$

- Let X be the random variable that is equal to the number of times the swap-operation is performed when running InsertionSort$(A[1 \ldots n])$. Determine the expected value $\mathbb{E}(X)$ of X.

6.19 Assume we have n balls and m boxes. We throw the balls independently and uniformly at random in the boxes. Thus, for $1 \leq k \leq n$ and $1 \leq i \leq m$,

$$\Pr(\text{ the } k\text{-th ball falls in the } i\text{-th box }) = 1/m.$$

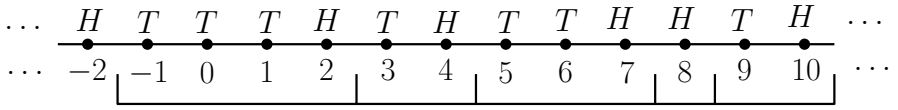
Define the following three random variables:

$$\begin{aligned} X &= \text{the number of boxes that do not contain any ball,} \\ Y &= \text{the number of boxes that contain at least one ball,} \\ Z &= \text{the number of boxes that contain exactly one ball.} \end{aligned}$$

- Determine the expected values $\mathbb{E}(X)$, $\mathbb{E}(Y)$, and $\mathbb{E}(Z)$.
- Assuming that $m = n$, determine the limits
  1. $\lim_{n\to\infty} \mathbb{E}(X)/n$,
  2. $\lim_{n\to\infty} \mathbb{E}(Y)/n$,
  3. $\lim_{n\to\infty} \mathbb{E}(Z)/n$.

  (Use the fact that $\lim_{n\to\infty}(1 - 1/n)^n = 1/e$.)

6.20 Let $0 < p < 1$ and consider a coin that comes up heads with probability p and tails with probability $1 - p$. For each integer n, let b_n be the outcome when flipping this coin; thus, $b_n \in \{H, T\}$. The values b_n partition the set of integers into intervals, where each interval is a maximal consecutive sequence of zero or more T's followed by one H:



- Consider the interval that contains the integer 0, and let X be its length. (In the example above, $X = 4$.) Determine the expected value $\mathbb{E}(X)$ of X.

  (*Hint:* Use the Linearity of Expectation. The answer is not $1/p$, which is the expected number of coin flips until the first H.)

6.21 Your friend Mick takes a permutation of $1, 2, \ldots, n$, stores it in boxes $B_1, B_2, \ldots, B_n$ (so that each box stores exactly one number), and then closes all boxes. You have no idea what the permutation is.

Mick opens the boxes $B_1, B_2, \ldots, B_n$, one after another. For each i with $1 \leq i \leq n$, just before opening box B_i, you have to guess which number is stored in it.

- Assume that, when you guess the number in box B_i, you do not remember the numbers stored in $B_1, B_2, \ldots, B_{i-1}$. Then, the only reasonable thing you can do is to take a random element in $\{1, 2, \ldots, n\}$ and guess that this random element is stored in B_i.

  Assume that you do this for each i with $1 \leq i \leq n$. Let X be the random variable whose value is equal to the number of times that your guess is correct. Compute the expected value $\mathbb{E}(X)$ of X.

- Now assume that your memory is perfect, so that, when you guess the number in box B_i, you know the numbers stored in $B_1, B_2, \ldots, B_{i-1}$.

  How would you make the n guesses such that the following is true: If Y is the random variable whose value is equal to the number of times that your guess is correct, then the expected value $\mathbb{E}(Y)$ of Y satisfies $\mathbb{E}(Y) = \Omega(\log n)$.

6.22 Recall that a permutation of the set $S = \{1, 2, \ldots, n\}$ is a bijection $f : S \to S$ (i.e., f is one-to-one and onto).

- Consider a fixed element k in the set S. How many permutations $f : S \to S$ are there for which $f(k) = k$?

- We choose a permutation $f : S \to S$ uniformly at random. Define the random variable
$$X = |\{k \in S : f(k) = k\}|.$$
Determine the expected value $\mathbb{E}(X)$.

6.23 Let d be the number of days in one year, and consider a group $P_1, P_2, \ldots, P_n$ of n people. Assume that each person has a random and independent birthday, i.e., any of the d days has a probability of $1/d$ of being the birthday of P_i, which is independent of the birthdays of the other people.

Define the random variable X to be the number of pairs $\{P_i, P_j\}$ of people that have the same birthday. Prove that
$$\mathbb{E}(X) = \frac{1}{d}\binom{n}{2}.$$

6.24 Michiel's Craft Beer Company (MCBC) sells n different brands of India Pale Ale (IPA). When you place an order, MCBC sends you one bottle of IPA, chosen uniformly at random from the n different brands, independently of previous orders.

Simon Pratt wants to try all different brands of IPA. He repeatedly places orders at MCBC (one bottle per order) until he has received at least one bottle of each brand.

Define the random variable X to be the total number of orders that Simon places. Determine the expected value $\mathbb{E}(X)$. (*Hint:* Use the Linearity of Expectation. If Simon has received exactly i different brands of IPA, how many orders does he expect to place until he receives a new brand?)

6.25 MCBC still sells n different brands of IPA. As in Exercise 6.24, when you place an order, MCBC sends you one bottle of IPA, chosen uniformly at random from the n different brands, independently of previous orders.

Simon Pratt places m orders at MCBC. Define the random variable X to be the total number of distinct brands that Simon receives. Determine the expected value $\mathbb{E}(X)$. (*Hint:* Use indicator random variables.)

6.26 You are given an array $A[0 \ldots n-1]$ of n numbers. Let D be the number of *distinct* numbers that occur in this array. For each i with $0 \leq i \leq n-1$, let N_i be the number of elements in the array that are equal to $A[i]$.

- Show that $D = \sum_{i=0}^{n-1} 1/N_i$.

Consider the following algorithm:

Algorithm EstimateD($A[1 \dots n]$)**:**

Step 1: Choose an integer k in $\{0, 1, 2, \dots, n-1\}$ uniformly at random, and let $a = A[k]$.
Step 2: Traverse the array and compute the number N_k of times that a occurs.
Step 3: Return the value $X = n/N_k$.

- Determine the expected value $\mathbb{E}(X)$ of the random variable X. (*Hint:* Use the definition of expected value, i.e., Definition 6.4.1.)

6.27 Consider a group of $n = 100,000$ people who play the following game. One of these people is chosen uniformly at random and wins $1,000$ dollars. All other persons do not win anything.

For each k with $1 \leq k \leq n$, let W_k be the random variable whose value is the amount of money that the k-th person wins.

- Determine the expected value $\mathbb{E}(W_k)$ of W_k.
- Determine $\max_{1 \leq k \leq n} \mathbb{E}(W_k)$.
- Define the random variable $W = \max_{1 \leq k \leq n} W_k$. Determine the expected value $\mathbb{E}(W)$ of W.
- Conclude that
$$\mathbb{E}\left(\max_{1 \leq k \leq n} W_k\right) \neq \max_{1 \leq k \leq n} \left(\mathbb{E}(W_k)\right).$$

6.28 Consider the sample space

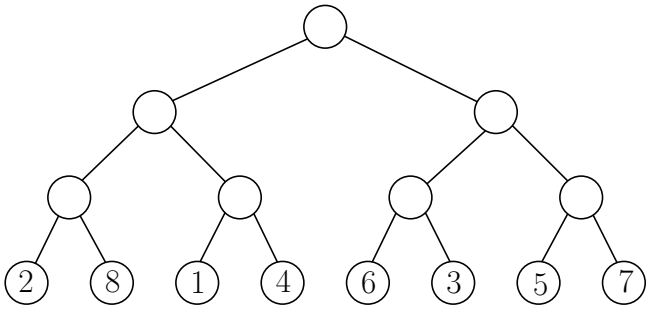
$$S = \{(123), (132), (213), (231), (312), (321), (111), (222), (333)\}.$$

We choose an element u from S uniformly at random. For each $i = 1, 2, 3$, let X_i be the random variable whose value is the i-th number in u. (For example, if $u = (312)$, then $X_1 = 3$, $X_2 = 1$, and $X_3 = 2$.) Let N be the random variable whose value is equal to that of X_2.

- Verify that $\Pr(X_i = k) = 1/3$ for any i and k with $1 \leq i \leq 3$ and $1 \leq k \leq 3$.

- Verify that X_1, X_2 and X_3 are pairwise independent.
- Verify that X_1, X_2 and X_3 are not mutually independent.
- Verify that $\sum_{i=1}^{\mathbb{E}(N)} \mathbb{E}(X_i) = 4$.
- Verify that $\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) \neq \sum_{i=1}^{\mathbb{E}(N)} \mathbb{E}(X_i)$.

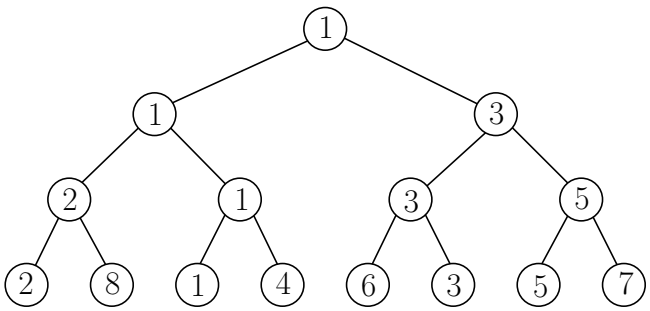
6.29 Let $n \geq 2$ be a power of two and consider a full binary tree with n leaves. Let $a_1, a_2, \ldots, a_n$ be a random permutation of the numbers $1, 2, \ldots, n$. Store this permutation at the leaves of the tree, in the order $a_1, a_2, \ldots, a_n$ from left to right. For example, if $n = 8$ and the permutation is $2, 8, 1, 4, 6, 3, 5, 7$, then we obtain the following tree:



We perform the following process on the tree:

- Visit the levels of the tree from bottom to top.
- At each level, take all pairs of consecutive nodes that have the same parent. For each such pair, compare the numbers stored at the two nodes, and store the smaller of these two numbers at the common parent.

For our example tree, we obtain the following tree:



It is clear that at the end of this process, the root stores the number 1. Define the random variable X to be the number that is not equal to 1 and that is stored at a child of the root. For our example tree, $X = 3$.

The following questions will lead you through a proof that $\mathbb{E}(X) \leq 3$.

- Prove that $2 \leq X \leq 1 + n/2$.
- Prove that the following is true for each k with $1 \leq k \leq n/2$: $X \geq k+1$ if and only if
  - either all numbers $1, 2, \ldots, k$ are stored in the left subtree of the root
  - or all numbers $1, 2, \ldots, k$ are stored in the right subtree of the root.
- Prove that for each k with $1 \leq k \leq n/2$,
$$\Pr(X \geq k+1) = \frac{2\binom{n/2}{k}k!(n-k)!}{n!}.$$
- Prove that for each k with $1 \leq k \leq n/2$,
$$\frac{2\binom{n/2}{k}k!(n-k)!}{n!} = \frac{(\frac{n}{2}-1)(\frac{n}{2}-2)(\frac{n}{2}-3)\cdots(\frac{n}{2}-k+1)}{(n-1)(n-2)(n-3)\cdots(n-k+1)}.$$
- Prove that for each i with $0 \leq i \leq n/2$,
$$\frac{\frac{n}{2}-i}{n-i} \leq 1/2.$$
- Prove that for each k with $1 \leq k \leq n/2$,
$$\Pr(X \geq k+1) \leq (1/2)^{k-1}.$$
- According to Exercise 6.8, we have
$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \Pr(X \geq k).$$
Prove that
$$\mathbb{E}(X) = \Pr(X \geq 1) + \sum_{k=1}^{n/2} \Pr(X \geq k+1).$$
- Prove that $\mathbb{E}(X) \leq 3$.

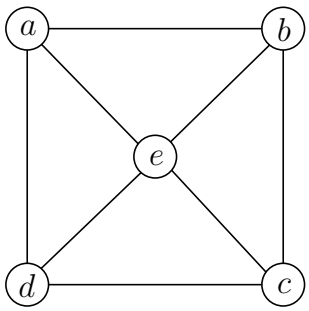
Chapter 7

The Probabilistic Method

The Probabilistic Method is a very powerful and surprising tool that uses probability theory to prove results in discrete mathematics. In this chapter, we will illustrate this method using several examples.

7.1 Large Bipartite Subgraphs

Recall that a *graph* is a pair $G = (V, E)$, where V is a finite set whose elements are called *vertices* and E is a set whose elements are unordered pairs of vertices. The elements of E are called *edges*. Assume we partition the vertex set V of G into two subsets A and B (thus, $A \cap B = \emptyset$ and $A \cup B = V$). We say that an edge of E is *between* A and B, if one vertex of this edge is in A and the other vertex is in B.



For example, in the graph above, let $A = \{a, d\}$ and $B = \{b, c, e\}$. Then four of the eight edges are between A and B, namely $\{a, b\}$, $\{a, e\}$, $\{d, c\}$, and $\{d, e\}$. Thus, the vertex set of this graph can be partitioned into two

subsets A and B, such that at least half of G's edges are between A and B. The following theorem states that this is true for any graph.

Theorem 7.1.1 *Let $G = (V, E)$ be a graph with m edges. The vertex set V of G can be partitioned into two subsets A and B such that the number of edges between A and B is at least $m/2$.*

Proof. Consider the following random process: Initialize $A = \emptyset$ and $B = \emptyset$. For each vertex u of G, flip a fair and independent coin. If the coin comes up heads, add u to A; otherwise, add u to B.

Define the random variable X to be the number of edges of G that are between A and B. We will determine the expected value $\mathbb{E}(X)$ of X.

Number the edges of G arbitrarily as $e_1, e_2, \ldots, e_m$. For each i with $1 \leq i \leq m$, define the indicator random variable

$$X_i = \begin{cases} 1 & \text{if } e \text{ is an edge between } A \text{ and } B, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$X = \sum_{i=1}^{m} X_i$$

and

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{m} X_i\right) \\ &= \sum_{i=1}^{m} \mathbb{E}\left(X_i\right) \\ &= \sum_{i=1}^{m} \Pr\left(X_i = 1\right). \end{aligned}$$

To determine $\Pr\left(X_i = 1\right)$, let e_i have vertices a and b. The following table shows the four possibilities for a and b:

$a \in A$, $b \in A$	$X_i = 0$
$a \in A$, $b \in B$	$X_i = 1$
$a \in B$, $b \in A$	$X_i = 1$
$a \in B$, $b \in B$	$X_i = 0$

Since $X_i = 1$ in two out of the four cases, we have

$$\Pr(X_i = 1) = 2/4 = 1/2,$$

and it follows that

$$\mathbb{E}(X) = \sum_{i=1}^{m} 1/2 = m/2.$$

Assume the claim in the theorem does not hold. Then, no matter how we partition the vertex set V into A and B, the number of edges between A and B will be less than $m/2$. In particular, the random variable X will always be less than $m/2$. But then, $\mathbb{E}(X) < m/2$ as well, contradicting that $\mathbb{E}(X) = m/2$. ∎

7.2 Ramsey Theory

We return to a problem we have seen in Section 1.1. Consider a complete graph with n vertices, in which each vertex represents a person. Any pair of vertices is connected by an edge. Such an edge is *solid* if the two persons representing the vertices of this edge are friends. If these persons are strangers, the edge is *dashed*. Consider a subset S of k vertices. We say that S is a *solid k-clique*, if any two vertices in S are connected by a solid edge. Thus, a solid k-clique represents a group of k mutual friends. If any two vertices of S are connected by a dashed edge, we say that S is a *dashed k-clique*; this represents a group of k mutual strangers.

In Section 1.1, we stated, without proof, Theorem 1.1.3. We repeat the statement of this theorem and use the Probabilistic Method to prove it.

Theorem 7.2.1 *Let $k \geq 3$ and $n \leq \lfloor 2^{k/2} \rfloor$ be integers. There exists a complete graph with n vertices, in which each edge is either solid or dashed, such that this graph does not contain a solid k-clique and does not contain a dashed k-clique.*

Proof. We denote the complete graph with n vertices by K_n. Consider the following random process: For each edge e of K_n, flip a fair and independent coin. If the coin comes up heads, make e a solid edge; otherwise, make e a dashed edge.

Define the event

$$A = \text{ "there is a solid } k\text{-clique or there is a dashed } k\text{-clique".}$$

We will prove below that $\Pr(A) < 1$. This will imply that $\Pr\left(\overline{A}\right) > 0$, i.e., the event

$$\overline{A} = \text{ "there is no solid } k\text{-clique and there is no dashed } k\text{-clique"}$$

has a positive probability. This, in turn, will imply that the statement in the theorem holds: If the statement would not hold, then $\Pr\left(\overline{A}\right)$ would be zero.

Thus, it remains to prove that $\Pr(A) < 1$. The vertex set of K_n has exactly $\binom{n}{k}$ many subsets of size k. We denote these subsets by V_i, $i = 1, 2, \ldots, \binom{n}{k}$. For each i with $1 \leq i \leq \binom{n}{k}$, define the event

$$A_i = \text{ "}V_i \text{ is a solid } k\text{-clique or a dashed } k\text{-clique".}$$

Since the event A_i occurs if and only if the edges joining the $\binom{k}{2}$ pairs of vertices of V_i are either all solid or all dashed, we have

$$\Pr(A_i) = \frac{2}{2^{\binom{k}{2}}}.$$

Since

$$A \text{ occurs if and only if } A_1 \vee A_2 \vee \cdots \vee A_{\binom{n}{k}} \text{ occurs,}$$

the Union Bound (i.e., Lemma 5.3.5) implies that

$$\begin{aligned}
\Pr(A) &= \Pr\left(A_1 \vee A_2 \vee \cdots \vee A_{\binom{n}{k}}\right) \\
&\leq \sum_{i=1}^{\binom{n}{k}} \Pr(A_i) \\
&= \sum_{i=1}^{\binom{n}{k}} \frac{2}{2^{\binom{k}{2}}} \\
&= \frac{2\binom{n}{k}}{2^{\binom{k}{2}}}.
\end{aligned}$$

If we can show that the quantity in the last line is less than one, then the proof is complete. We have

$$\begin{aligned}\frac{2\binom{n}{k}}{2^{\binom{k}{2}}} &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\cdot\frac{2}{2^{(k^2-k)/2}} \\ &\leq \frac{n^k}{k!}\cdot\frac{2^{1+k/2}}{2^{k^2/2}}.\end{aligned}$$

Since $n \leq \lfloor 2^{k/2} \rfloor \leq 2^{k/2}$, we get

$$\begin{aligned}\frac{2\binom{n}{k}}{2^{\binom{k}{2}}} &\leq \frac{(2^{k/2})^k}{k!}\cdot\frac{2^{1+k/2}}{2^{k^2/2}} \\ &= \frac{2^{1+k/2}}{k!}.\end{aligned}$$

By Exercise 2.8, we have $k! > 2^{1+k/2}$ for $k \geq 3$. Thus, we conclude that

$$\frac{2\binom{n}{k}}{2^{\binom{k}{2}}} < 1.$$

∎

Take, for example, $k = 20$ and $n = 1024$. Theorem 7.2.1 states that there exists a group of 1024 people that does not contain a subgroup of 20 mutual friends and does not contain a subgroup of 20 mutual strangers. In fact, the proof shows more: Consider a group of 1024 people such that any two are friends with probability 1/2, and strangers with probability 1/2. The above proof shows that $\Pr(A)$, i.e., the probability that there is a subgroup of 20 mutual friends or there is a subgroup of 20 mutual strangers, satisfies

$$\Pr(A) \leq \frac{2^{1+k/2}}{k!} = \frac{2^{11}}{20!}.$$

Therefore, with probability at least

$$1 - \frac{2^{11}}{20!} = 0.99999999999999158,$$

(there are 15 nines) this group does not contain a subgroup of 20 mutual friends and does not contain a subgroup of 20 mutual strangers.

7.3 Sperner's Theorem

In Section 1.2, we considered the following problem. Let S be a set of size n and consider a sequence $S_1, S_2, \ldots, S_m$ of m subsets of S, such that for all i and j with $i \neq j$,

$$S_i \not\subseteq S_j \text{ and } S_j \not\subseteq S_i. \tag{7.1}$$

What is the largest possible value of m for which such a sequence exists?

The sequence consisting of all subsets of S having size $\lfloor n/2 \rfloor$ satisfies (7.1). This sequence has length $m = \binom{n}{\lfloor n/2 \rfloor}$. In Section 1.2, we stated, without proof, that this is the largest possible value of m; see Theorem 1.2.1. After stating this theorem again, we will prove it using the Probabilistic Method.

Theorem 7.3.1 (Sperner) *Let $n \geq 1$ be an integer and let S be a set with n elements. Let $S_1, S_2, \ldots, S_m$ be a sequence of m subsets of S, such that for all i and j with $i \neq j$,*

$$S_i \not\subseteq S_j \text{ and } S_j \not\subseteq S_i.$$

Then

$$m \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Proof. We assume that none of the subsets in the sequence $S_1, S_2, \ldots, S_m$ is empty, because otherwise, m must be equal to 1, in which case the theorem clearly holds.

We assume that $S = \{1, 2, \ldots, n\}$. Recall that a permutation of S is a bijection $f : S \to S$. We choose a uniformly random permutation f of S; thus, each permutation has probability $1/n!$ of being chosen. Define the following sequence of subsets $A_1, A_2, \ldots, A_n$ of S: For $j = 1, 2, \ldots, n$,

$$A_j = \{f(1), f(2), \ldots, f(j)\}.$$

For example, if $n = 4$ and the permutation f is given by $f(1) = 3$, $f(2) = 1$, $f(3) = 4$, and $f(4) = 2$, then

$$\begin{aligned} A_1 &= \{3\}, \\ A_2 &= \{1, 3\}, \\ A_3 &= \{1, 3, 4\}, \\ A_4 &= \{1, 2, 3, 4\}. \end{aligned}$$

Observe that the subsets $A_1, A_2, \ldots, A_n$ are random subsets of S, because f was randomly chosen.

Consider a subset S_i in the statement of the theorem. We say that S_i *occurs* in the sequence $A_1, A_2, \ldots, A_n$ if there is an index j such that $S_i = A_j$.

Define the random variable X to be the number of subsets in the sequence $S_1, S_2, \ldots, S_m$ that occur in $A_1, A_2, \ldots, A_n$. Since the subsets $A_1, A_2, \ldots, A_n$ are properly nested, i.e.,

$$A_1 \subset A_2 \subset \cdots \subset A_n,$$

the assumption in the theorem implies that X is either 0 or 1. It follows that the expected value of X satisfies

$$\mathbb{E}(X) \leq 1.$$

Below, we will determine the exact value of $\mathbb{E}(X)$.

For each i with $1 \leq i \leq m$, define the indicator random variable

$$X_i = \begin{cases} 1 & \text{if } S_i \text{ occurs in the sequence } A_1, A_2, \ldots, A_n, \\ 0 & \text{otherwise.} \end{cases}$$

Let k denote the size of the subset S_i, i.e., $k = |S_i|$. Then

$$X_i = 1 \text{ if and only if } S_i = A_k.$$

Since $A_k = \{f(1), f(2), \ldots, f(k)\}$, $X_i = 1$ if and only if the first k values in the permutation f form a permutation of the subset S_i:

$$\underbrace{\boxed{f(1), f(2), \ldots, f(k)}}_{\text{permutation of } S_i} \boxed{f(k+1), f(k+2), \ldots, f(n)}$$

The Product Rule of Section 3.1 shows that there are $k!(n-k)!$ many permutations f of S that have this property. Therefore, since f is a random permutation of S, we have

$$\begin{aligned} \mathbb{E}(X_i) &= \Pr(X_i = 1) \\ &= \frac{k!(n-k)!}{n!} \\ &= \frac{1}{\binom{n}{k}} \\ &= \frac{1}{\binom{n}{|S_i|}}. \end{aligned}$$

Thus, since

$$X = \sum_{i=1}^{m} X_i,$$

we get

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left(\sum_{i=1}^{m} X_i\right) \\ &= \sum_{i=1}^{m} \mathbb{E}\left(X_i\right) \\ &= \sum_{i=1}^{m} \frac{1}{\binom{n}{|S_i|}}. \end{aligned}$$

If we combine this with our upper bound $\mathbb{E}(X) \leq 1$, we get

$$\sum_{i=1}^{m} \frac{1}{\binom{n}{|S_i|}} \leq 1.$$

For a fixed value of n, the binomial coefficient $\binom{n}{k}$ is maximized when $k = \lfloor n/2 \rfloor$; i.e., the largest value in the n-th row of Pascal's Triangle (see Section 3.8) is in the middle. Thus,

$$\binom{n}{|S_i|} \leq \binom{n}{\lfloor n/2 \rfloor},$$

implying that

$$\begin{aligned} 1 &\geq \sum_{i=1}^{m} \frac{1}{\binom{n}{|S_i|}} \\ &\geq \sum_{i=1}^{m} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \\ &= \frac{m}{\binom{n}{\lfloor n/2 \rfloor}}. \end{aligned}$$

We conclude that

$$m \leq \binom{n}{\lfloor n/2 \rfloor}.$$

■

7.4 Planar Graphs and the Crossing Lemma

Consider a graph $G = (V, E)$. Any one-to-one function $f : V \to \mathbb{R}^2$ gives an *embedding* of G:

1. Each vertex a of V is drawn as the point $f(a)$ in the plane.

2. Each edge $\{a, b\}$ of E is drawn as the straight-line segment $f(a)f(b)$ between the points $f(a)$ and $f(b)$.

Besides the function f being one-to-one, we assume that it satisfies the following three properties:

1. For any two edges $\{a, b\}$ and $\{a', b'\}$ of E, the intersection of the line segments $f(a)f(b)$ and $f(a')f(b')$ is empty or consists of exactly one point.

2. For any edge $\{a, b\}$ in E and any vertex c in V, the point $f(c)$ is not in the interior of the line segment $f(a)f(b)$.

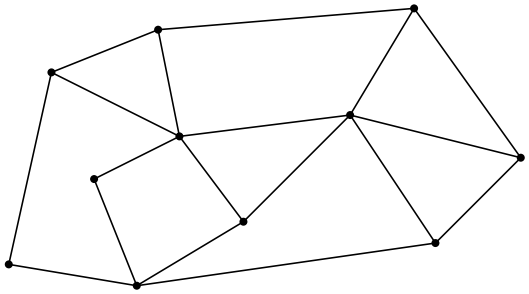
3. For any three edges $\{a, b\}$, $\{a', b'\}$, and $\{a'', b''\}$ of E, the line segments $f(a)f(b)$, $f(a')f(b')$, and $f(a'')f(b'')$ do not have a point in common that is in the interior of any of these line segments.

For simplicity, we do not distinguish any more between a graph and its embedding. That is, a vertex a refers to both an element of V and the point in the plane that represents a. Similarly, an edge refers to both an element of E and the line segment that represents it.

7.4.1 Planar Graphs

Definition 7.4.1 An embedding of a graph $G = (V, E)$ is called *plane*, if no two edges of E intersect, except possibly at their endpoints. A graph G is called *planar* if there is a plane embedding of G.

Consider a plane embedding of a planar graph G. Again for simplicity, we denote this embedding by G. This embedding consists of vertices, edges, and faces (one of them being the unbounded face). For example, in the following plane embedding, there are 11 vertices, 18 edges, and 9 faces.



In the rest of this section, we will use the following notation:

- G denotes a plane embedding of a planar graph.
- v denotes the number of vertices of G.
- e denotes the number of edges of G.
- f denotes the number of faces in the embedding of G.

How many edges can G have? Since G has v vertices, we obviously have $e \leq \binom{v}{2} = \Theta(v^2)$, an upper bound which holds for any graph with v vertices. Since our graph G is planar, we expect a much smaller upper bound on e: If G has $\Theta(v^2)$ edges, then it seems to be impossible to draw G without edge crossings. Below, we will prove that e is, in fact, at most linear in v. The proof will use Euler's Theorem for planar graphs:

Theorem 7.4.2 (Euler) *Consider any plane embedding of a planar graph G. Let v, e, and f be the number of vertices, edges, and faces of this embedding, respectively. Moreover, let c be the number of connected components of G. Then*

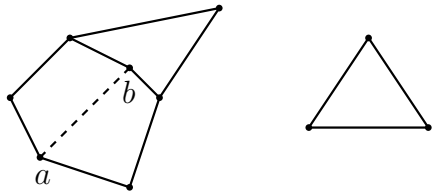
$$v - e + f = c + 1. \tag{7.2}$$

Proof. The idea of the proof is as follows. We start by removing all edges from G (but keep all vertices), and show that (7.2) holds. Then we add back the edges of G, one by one, and show that (7.2) remains valid throughout this process.

After having removed all edges, we have $e = 0$ and the embedding consists of a collection of v points. Since $f = 1$ and $c = v$, the relation $v-e+f = c+1$ holds.

Assume the relation $v - e + f = c + 1$ holds and consider what happens when we add an edge ab. There are two possible cases.

Case 1: Before adding the edge ab, the vertices a and b belong to the same connected component.

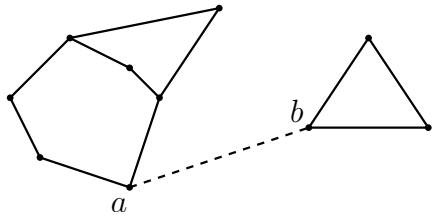


When adding the edge ab,

- the number v of vertices does not change,
- the number e of edges increases by one,
- the number f of faces increases by one (because the edge ab splits one face into two),
- the number c of connected components does not change.

It follows that the relation $v - e + f = c + 1$ still holds after ab has been added.

Case 2: Before adding the edge ab, the vertices a and b belong to different connected components.



When adding the edge ab,

- the number v of vertices does not change,
- the number e of edges increases by one,
- the number f of faces does not change,
- the number c of connected components decreases by one.

It again follows that the relation $v - e + f = c + 1$ still holds after ab has been added. ∎

Usually, Euler's Theorem is stated for connected planar graphs, i.e., planar graphs for which $c = 1$:

Theorem 7.4.3 (Euler) *Consider any plane embedding of a connected planar graph G. If v, e, and f denote the number of vertices, edges, and faces of this embedding, respectively, then*

$$v - e + f = 2.$$

We now show how to use Euler's Theorem to prove an upper bound on the number of edges and faces of any connected planar graph:

Theorem 7.4.4 *Let G be any plane embedding of a connected planar graph with $v \geq 3$ vertices. Then*

1. *G has at most $3v - 6$ edges and*

2. *this embedding has at most $2v - 4$ faces.*

Proof. As before, let e and f denote the number of edges and faces of G, respectively. If $v = 3$, then $e \leq 3$ and $f \leq 2$. Hence, in this case, we have $e \leq 3v - 6$ and $f \leq 2v - 4$.

Assume that $v \geq 4$. We number the faces of G arbitrarily from 1 to f. For each i with $1 \leq i \leq f$, let m_i denote the number of edges on the i-th face of G. Since each edge lies on the boundary of at most two faces, the summation $\sum_{i=1}^{f} m_i$ counts each edge at most twice. Thus,

$$\sum_{i=1}^{f} m_i \leq 2e.$$

On the other hand, since G is connected and $v \geq 4$, each face has at least three edges on its boundary, i.e., $m_i \geq 3$. It follows that

$$\sum_{i=1}^{f} m_i \geq 3f.$$

Combining these two inequalities implies that $3f \leq 2e$, which we rewrite as

$$f \leq 2e/3.$$

Using Euler's formula (with $c = 1$, because G is connected), we obtain

$$e = v + f - 2 \leq v + 2e/3 - 2,$$

which is equivalent to

$$e \leq 3v - 6.$$

We also obtain

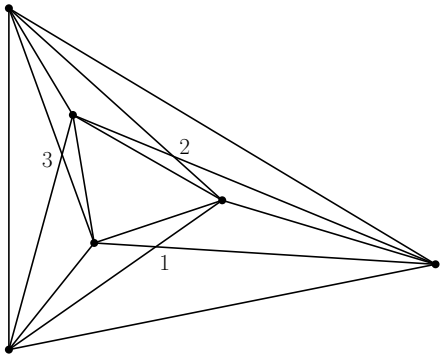
$$f \leq 2e/3 \leq 2(3v - 6)/3 = 2v - 4.$$

This completes the proof. ∎

7.4.2 The Crossing Number of a Graph

Consider an embedding of a graph $G = (V, E)$. We say that two distinct edges of E *cross*, if their interiors have a point in common. In this case, we call this common point a *crossing*.

The example below shows an embedding of the complete graph K_6 on six vertices, which are denoted by black dots. In this embedding, there are three crossings, which are numbered 1, 2, and 3.



Definition 7.4.5 The *crossing number* $cr(G)$ of a graph G is defined to be the minimum number of crossings in any embedding of G.

Thus, a graph G is planar if and only if $cr(G) = 0$. The example above shows that $cr(K_6) \leq 3$.

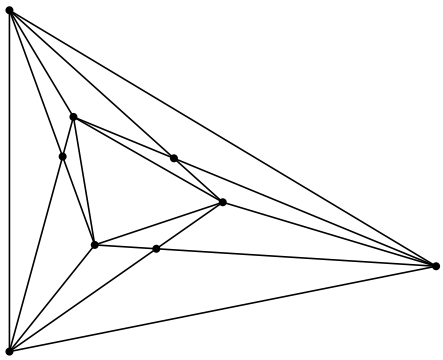
In the rest of this section, we consider the following problem: Given a graph G with v vertices and e edges, can we prove good bounds, in terms of v and e, on the crossing number $cr(G)$ of G?

A simple lower bound on the crossing number

Let G be any graph with $v \geq 3$ vertices and e edges. Consider an embedding of G having $cr(G)$ crossings; hence, this embedding is the "best" one.

We "make" G planar, by defining all crossings to be vertices. That is, let H be the graph whose vertex set is the union of the vertex set of G and the set of all crossings in the embedding. Edges of G are cut by the crossings into smaller edges, which are edges in the graph H.

The figure below shows the planar version of the embedding of K_6 that we saw before. This new graph has 9 vertices and 21 edges.



We make the following observations:

- The graph H is planar, because it is embedded without any crossings.
- The graph H has $v + cr(G)$ vertices.
- How many edges does H have? Any crossing in G is the intersection of exactly two edges of G; these two edges contribute four edges to H. Hence, for any crossing in G, the number of edges in H increases by two. It follows that H has $e + 2 \cdot cr(G)$ edges.

Since H is planar, we know from Theorem 7.4.4 that the number of its edges is bounded from above by three times the number of its vertices minus six, i.e.,

$$e + 2 \cdot cr(G) \leq 3(v + cr(G)) - 6.$$

By rewriting this inequality, we obtain the following result:

Theorem 7.4.6 *For any graph G with $v \geq 3$ vertices and e edges, we have*

$$cr(G) \geq e - 3v + 6.$$

For example, consider the complete graph K_n on n vertices, where $n \geq 3$. Since this graph has $\binom{n}{2}$ edges, we obtain

$$cr\,(K_n) \geq \binom{n}{2} - 3n + 6 = \frac{1}{2}n^2 - \frac{7}{2}n + 6. \tag{7.3}$$

For $n = 6$, we get $cr(K_6) \geq 3$. Since we have seen before that $cr(K_6) \leq 3$, it follows that $cr(K_6) = 3$.

Since K_n has $\binom{n}{2}$ edges and any two of them cross at most once, we have the following obvious upper bound on the crossing number of K_n:

$$cr\,(K_n) \leq \binom{\binom{n}{2}}{2} = O(n^4). \tag{7.4}$$

(Of course (7.4) holds for any graph with n vertices.)

To conclude this subsection, (7.3) gives an n^2-lower bound, whereas (7.4) gives an n^4-upper bound on the crossing number of K_n. In the next section, we will determine the true asymptotic behavior of $cr(K_n)$.

A better lower bound on the crossing number

As before, let G be any graph with $v \geq 3$ vertices and e edges. Again we consider an embedding of G having $cr(G)$ crossings. In the rest of this subsection, we will use the Probabilistic Method to prove a lower bound on the crossing number of G.

We choose a real number p such that $0 < p \leq 1$. Consider a coin that comes up heads with probability p and comes up tails with probability $1 - p$. Let G_p be the random subgraph of G, that is obtained as follows.

- For each vertex a of G, flip the coin (independently of the other coin flips) and add a as a vertex to G_p if and only if the coin comes up heads.

- Each edge ab of G appears as an edge in G_p if and only if both a and b are vertices of G_p.

Recall that we fixed the embedding of G. As a result, this random process gives us an embedding of G_p (which may not be the best one in terms of the number of crossings.)

We denote the number of vertices, edges, and crossings in the embedding of G_p by v_p, e_p, and x_p, respectively. Observe that these three quantities are random variables.

It follows from Theorem 7.4.6 that

$$cr(G_p) - e_p + 3v_p \geq 6,$$

provided that $v_p \geq 3$. This implies that

$$cr(G_p) - e_p + 3v_p \geq 0,$$

for any value of v_p that results from our random choices.

Since $cr(G_p) \leq x_p$, it follows that

$$x_p - e_p + 3v_p \geq 0.$$

The left-hand side is a random variable, which is always non-negative, no matter what graph G_p results from our random choices. Therefore, its expected value is also non-negative, i.e.,

$$\mathbb{E}\left(x_p - e_p + 3v_p\right) \geq 0.$$

Using the Linearity of Expectation (i.e., Theorem 6.5.2), we get

$$\mathbb{E}(x_p) - \mathbb{E}(e_p) + 3 \cdot \mathbb{E}(v_p) \geq 0. \tag{7.5}$$

We are now going to compute these three expected values separately.

The random variable v_p is equal to the number of successes in v independent trials, each one having success probability p. In other words, v_p has a binomial distribution with parameters v and p, and, therefore, by Theorem 6.7.2,

$$\mathbb{E}(v_p) = pv.$$

To compute $\mathbb{E}(e_p)$, we number the edges of G arbitrarily from 1 to e. For each i with $1 \leq i \leq e$, define X_i to be the indicator random variable with value

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th edge is an edge in } G_p, \\ 0 & \text{otherwise.} \end{cases}$$

Since an edge of G is in G_p if and only if both its vertices are in G_p, it follows that

$$\mathbb{E}(X_i) = \Pr(X_i = 1) = p^2.$$

Then, since $e_p = \sum_{i=1}^{e} X_i$, we get

$$\mathbb{E}(e_p) = \mathbb{E}\left(\sum_{i=1}^{e} X_i\right) = \sum_{i=1}^{e} \mathbb{E}(X_i) = \sum_{i=1}^{e} p^2 = p^2 e.$$

Finally, we compute the expected value of the random variable x_p. Number the crossings in the embedding of G arbitrarily from 1 to $cr(G)$. For each i with $1 \leq i \leq cr(G)$, define Y_i to be the indicator random variable with value

$$Y_i = \begin{cases} 1 & \text{if the } i\text{-th crossing is a crossing in } G_p, \\ 0 & \text{otherwise.} \end{cases}$$

Let ab and cd be the edges of G that cross in the i-th crossing[1]. This crossing appears as a crossing in G_p if and only if both ab and cd are edges in G_p. Since the points a, b, c, and d are pairwise distinct, it follows that the i-th crossing of G appears as a crossing in G_p with probability p^4. Thus,

$$\mathbb{E}(Y_i) = \Pr(Y_i = 1) = p^4.$$

Since $x_p = \sum_{i=1}^{cr(G)} Y_i$, it follows that

$$\mathbb{E}(x_p) = \mathbb{E}\left(\sum_{i=1}^{cr(G)} Y_i\right) = \sum_{i=1}^{cr(G)} \mathbb{E}(Y_i) = \sum_{i=1}^{cr(G)} p^4 = p^4 \cdot cr(G).$$

Substituting the three expected values into (7.5), we get

$$p^4 \cdot cr(G) - p^2 e + 3 \cdot pv \geq 0,$$

which we rewrite as

$$cr(G) \geq \frac{p^2 e - 3pv}{p^4}. \tag{7.6}$$

Observe that this inequality holds for *any* real number p with $0 < p \leq 1$.

If we assume that $e \geq 4v$, and take $p = 4v/e$ (so that $0 < p \leq 1$), then we obtain a new lower bound on the crossing number:

[1]By our definition of embedding, see Section 7.4.1, there are exactly two edges that determine the i-th crossing.

Theorem 7.4.7 (Crossing Lemma) *Let G be any graph with v vertices and e edges. If $e \geq 4v$, then*

$$cr(G) \geq \frac{1}{64}\frac{e^3}{v^2}.$$

Applying this lower bound to the complete graph K_n gives $cr(K_n) = \Omega(n^4)$. This lower bound is much better than the quadratic lower bound in (7.3) and it matches the upper bound in (7.4). Hence, we have shown that $cr(K_n) = \Theta(n^4)$.

Remark 7.4.8 Let n be a very large integer and consider the complete graph K_n with $v = n$ vertices and $e = \binom{n}{2}$ edges. Let us see what happens if we repeat the proof for this graph. We choose a random subgraph G_p of K_n, where $p = 4v/e = 8/(n-1)$. The expected number of vertices in G_p is equal to pn, which is approximately equal to 8. Thus, the random graph G_p is, expected, extremely small. Then we apply the *weak* lower bound of Theorem 7.4.6 to this, again expected, *extremely small* graph. The result is a proof that in any embedding of the *huge* graph K_n, there are $\Omega(n^4)$ crossings!

7.5 Exercises

7.1 Prove that Theorem 7.4.4 also holds if G is not connected.

7.2 Let K_5 be the complete graph on 5 vertices. In this graph, each pair of vertices is connected by an edge. Prove that K_5 is not planar.

7.3 Let G be any embedding of a connected planar graph with $v \geq 4$ vertices. Assume that this embedding has no triangles, i.e., there are no three vertices a, b, and c, such that ab, bc, and ac are edges of G.

- Prove that G has at most $2v - 4$ edges.
- Let $K_{3,3}$ be the complete bipartite graph on 6 vertices. The vertex set of this graph consists of two sets A and B, both of size three, and each vertex of A is connected by an edge to each vertex of B. Prove that $K_{3,3}$ is not planar.

7.4 Consider the numbers R_n that were defined in Section 4.6. In Section 4.6.1, we proved that $R_n = O(n^8)$. Prove that $R_n = O(n^4)$.

7.5 Let n be a sufficiently large positive integer and consider the complete graph K_n. This graph has vertex set $V = \{1, 2, \ldots, n\}$, and each pair of distinct vertices is connected by an undirected edge. (Thus, K_n has $\binom{n}{2}$ edges.)

Let $\vec{K}_n$ be the directed graph obtained by making each edge $\{i, j\}$ of K_n a directed edge; thus, in $\vec{K}_n$, this edge either occurs as the directed edge (i, j) from i to j or as the directed edge (j, i) from j to i.

We say that three pairwise distinct vertices i, j, and k define a *directed triangle* in $\vec{K}_n$, if

- (i, j), (j, k), and (k, i) are edges in $\vec{K}_n$ or
- (i, k), (k, j), and (j, i) are edges in $\vec{K}_n$.

Prove that there exists a way to direct the edges of K_n, such that the number of directed triangles in $\vec{K}_n$ is at least $\frac{1}{4}\binom{n}{3}$.