Krzysztof Wojtas

Solutions to exercises and problems

from

Introduction to Algorithms
Fourth Edition

by Thomas H. Cormen
Charles E. Leiserson
Ronald L. Rivest
Clifford Stein

June 23, 2023

Disclaimer

This document is incomplete and was pre-released early to make easier access to the material for beta readers. Please be aware that the solutions in this document may contain bugs, be untested or badly formatted. Until the material reaches a mature state, I hope to receive feedback via the project's GitHub. There you can also track progress and read about the project's history and future:

`https://github.com/wojtask/clrs4e-solutions`

Contents

Part VIII Appendix: Mathematical Background

A Summations

A.1 Summation formulas and properties

A.1-1 Let $g_1, g_2, \ldots, g_n$ be functions, so that $g_k(i) = O(f_k(i))$ for each $k = 1, 2, \ldots, n$. From the definition of O-notation, there exist positive constants $c_1, c_2, \ldots, c_n$ and i_0, such that $g_k(i) \le c_k f_k(i)$ for all $i \ge i_0$. Let $c = \max\{c_k : 1 \le k \le n\}$. For $i \ge i_0$ we have

$$\begin{aligned}\sum_{k=1}^{n} g_k(i) &\le \sum_{k=1}^{n} c_k f_k(i) \\ &\le c \sum_{k=1}^{n} f_k(i) \\ &= O\left(\sum_{k=1}^{n} f_k(i)\right).\end{aligned}$$

A.1-2

$$\begin{aligned}\sum_{k=1}^{n}(2k-1) &= 2\sum_{k=1}^{n} k - \sum_{k=1}^{n} 1 && \text{(by the linearity property)} \\ &= \frac{2n(n+1)}{2} - n && \text{(by equation (A.1))} \\ &= n^2\end{aligned}$$

A.1-3 If we let $x = 10$ in equation (A.6), we obtain:

$$\begin{aligned} 111{,}111{,}111 &= \sum_{k=0}^{8} 10^k \\ &= \frac{10^9 - 1}{10 - 1}. \end{aligned}$$

A.1-4 Each subsequent term is the previous term multiplied by $-1/2$, so letting $x = -1/2$ in equation (A.7) gives us the value of the series:

$$\begin{aligned} \sum_{k=0}^{\infty} (-1/2)^k &= \frac{1}{1 - (-1/2)} \\ &= 2/3. \end{aligned}$$

A.1-5 For the upper bound, we bound k by n:

$$\begin{aligned} \sum_{k=1}^{n} k^c &\le \sum_{k=1}^{n} n^c \\ &= n \cdot n^c \\ &= O(n^{c+1}). \end{aligned}$$

To get the lower bound, we drop around half of the smallest terms and bound the remaining ones:

$$\begin{aligned} \sum_{k=1}^{n} k^c &\ge \sum_{k=\lceil n/2 \rceil}^{n} k^c \\ &\ge \sum_{k=\lceil n/2 \rceil}^{n} \lceil n/2 \rceil^c \\ &= (n - \lceil n/2 \rceil + 1)\lceil n/2 \rceil^c \\ &= (\lfloor n/2 \rfloor + 1)\lceil n/2 \rceil^c \\ &\ge \lceil n/2 \rceil^{c+1} \\ &\ge n^{c+1}/2^{c+1} \\ &= \Omega(n^{c+1}) \qquad \text{(since } 1/2^{c+1} \text{ is a constant).} \end{aligned}$$

By applying Theorem 3.1, we obtain the tight bound.

A.1-6 Let's differentiate both sides of equation (A.11):

$$\begin{aligned}\sum_{k=0}^{\infty} k \cdot k x^{k-1} &= \frac{(1-x)^2 + 2x(1-x)}{(1-x)^4} \\ &= \frac{(1-x) + 2x}{(1-x)^3} \\ &= \frac{1+x}{(1-x)^3}.\end{aligned}$$

To get the desired result, we multiply both sides of the equation by x.

A.1-7 The asymptotic upper bound:

$$\begin{aligned}\sum_{k=1}^{n} \sqrt{k \lg k} &\le \sum_{k=1}^{n} \sqrt{n \lg n} \\ &= n \cdot \sqrt{n \lg n} \\ &= O(n^{3/2} \lg^{1/2} n).\end{aligned}$$

For the asymptotic lower bound, assume for convenience that n is even and at least 4. Then:

$$\begin{aligned}\sum_{k=1}^{n} \sqrt{k \lg k} &> \sum_{k=n/2+1}^{n} \sqrt{k \lg k} \\ &> \sum_{k=n/2+1}^{n} \sqrt{(n/2) \lg(n/2)} \\ &= (n/2) \sqrt{(n/2)(\lg n - 1)} \\ &\ge (n/2) \sqrt{(n/2)(\lg n)/2} \\ &= (n^{3/2} \lg^{1/2} n)/4 \\ &= \Omega(n^{3/2} \lg^{1/2} n).\end{aligned}$$

The asymptotically tight bound follows from Theorem 3.1.

A.1-8

$$\begin{aligned}\sum_{k=1}^{n}\frac{1}{2k-1} &< 1+\sum_{k=2}^{n+1}\frac{1}{2k-2}\\ &= 1+\sum_{k=1}^{n}\frac{1}{2k}\\ &= 1+\frac{H_n}{2}\\ &= 1+\frac{\ln n+O(1)}{2} && \text{(by equation (A.9))}\\ &= \ln\sqrt{n}+O(1)\end{aligned}$$

A.1-9

$$\begin{aligned}\sum_{k=0}^{\infty}\frac{k-1}{2^k} &= \sum_{k=0}^{\infty}\frac{k}{2^k}-\sum_{k=0}^{\infty}\frac{1}{2^k}\\ &= \sum_{k=0}^{\infty}k(1/2)^k-\sum_{k=0}^{\infty}(1/2)^k\\ &= \frac{1/2}{(1-1/2)^2}-\frac{1}{1-1/2} && \text{(by equations (A.11) and (A.7))}\\ &= 0\end{aligned}$$

A.1-10

$$\begin{aligned}\sum_{k=1}^{\infty}(2k+1)x^{2k} &= 2\sum_{k=0}^{\infty}k(x^2)^k+\sum_{k=0}^{\infty}(x^2)^k-1\\ &= \frac{2x^2}{(1-x^2)^2}+\frac{1}{1-x^2}-1 && \text{(by equations (A.11) and (A.7))}\\ &= \frac{x^2(3-x^2)}{(1-x^2)^2}\end{aligned}$$

A.1-11

$$\begin{aligned}\prod_{k=2}^{n}\left(1-\frac{1}{k^2}\right) &= \prod_{k=2}^{n}\frac{k^2-1}{k^2}\\ &= \frac{\prod_{k=2}^{n}(k^2-1)}{\prod_{k=2}^{n}k^2}\\ &= \frac{\prod_{k=2}^{n}(k-1)\cdot\prod_{k=2}^{n}(k+1)}{\left(\prod_{k=1}^{n}k\right)^2}\\ &= \frac{\prod_{k=1}^{n-1}k\cdot\prod_{k=3}^{n+1}k}{(n!)^2}\\ &= \frac{(n-1)!\cdot\frac{(n+1)!}{2}}{(n!)^2}\\ &= \frac{n+1}{2n}\end{aligned}$$

A.2 Bounding summations

A.2-1 Using the properties of telescoping series, we get

$$\begin{aligned}\sum_{k=1}^{n}\frac{1}{k^2} &\le 1+\sum_{k=2}^{n}\frac{1}{k(k-1)}\\ &= 1+\sum_{k=2}^{n}\left(\frac{1}{k-1}-\frac{1}{k}\right)\\ &= 1+1-\frac{1}{n}\\ &< 2.\end{aligned}$$

A.2-2

$$\begin{aligned}
\sum_{k=0}^{\lfloor \lg n \rfloor} \left\lceil \frac{n}{2^k} \right\rceil &< \sum_{k=0}^{\lfloor \lg n \rfloor} \left(\frac{n}{2^k} + 1 \right) && \text{(by inequality (3.2))} \\
&\le \lg n + 1 + \sum_{k=0}^{\lfloor \lg n \rfloor} \frac{n}{2^k} \\
&< \lg n + 1 + \sum_{k=0}^{\infty} \frac{n}{2^k} \\
&= \lg n + 1 + \frac{n}{1 - 1/2} && \text{(by equation (A.7))} \\
&= \lg n + 1 + 2n \\
&= O(n)
\end{aligned}$$

A.2-3

Proceeding similarly as when searching for an upper bound on the nth harmonic number, we split the range 1 to n into $\lfloor \lg n \rfloor$ pieces and lower-bound the contribution of each piece by $1/2$:

$$\begin{aligned}
\sum_{k=1}^{n} \frac{1}{k} &\ge \sum_{i=0}^{\lfloor \lg n \rfloor - 1} \sum_{j=0}^{2^i - 1} \frac{1}{2^i + j} \\
&\ge \sum_{i=0}^{\lfloor \lg n \rfloor - 1} \sum_{j=0}^{2^i - 1} \frac{1}{2^{i+1}} \\
&= \sum_{i=0}^{\lfloor \lg n \rfloor - 1} (1/2) \\
&= \lfloor \lg n \rfloor / 2 \\
&= \Omega(\lg n).
\end{aligned}$$

A.2-4

The function $f(k) = k^3$ is monotonically increasing, so we approximate the summation using inequality (A.18):

$$\int_0^n x^3 \, dx \le \sum_{k=1}^{n} k^3 \le \int_1^{n+1} x^3 \, dx.$$

The integral approximation gives the upper bound

$$\int_0^n x^3 \, dx = \frac{n^4}{4},$$

and the lower bound

$$\int_1^{n+1} x^3\,dx = \frac{(n+1)^4 - 1}{4}.$$

A.2-5 Applying inequality (A.19) to $\sum_{k=1}^{n} 1/k$ leads to an improper integral:

$$\begin{aligned}\sum_{k=1}^{n} \frac{1}{k} &\le \int_0^n \frac{dx}{x} \\ &= \lim_{a \to 0^+} \int_a^n \frac{dx}{x} \\ &= \infty.\end{aligned}$$

As a result, we don't get any useful information about the upper bound on the summation. By rewriting it as $1 + \sum_{k=2}^{n} 1/k$, we can apply formula (A.19) to the second term and upper-bound the initial summation by $\ln n + 1$.

Problems

A-1 ***Bounding summations***

In order to determine the asymptotic tight bounds on each summation, we will find their asymptotic upper bound and asymptotic lower bound by substituting each term in the summations by appropriate values. Since the bounds on the summations don't depend on the parity of n, for simplicity we will assume that n is even.

a. An asymptotic upper bound:

$$\begin{aligned}\sum_{k=1}^{n} k^r &\le \sum_{k=1}^{n} n^r \\ &= n \cdot n^r \\ &= O(n^{r+1}).\end{aligned}$$

An asymptotic lower bound:

$$\begin{aligned}\sum_{k=1}^{n} k^r &\geq \sum_{k=n/2+1}^{n} k^r \\ &\geq \sum_{k=n/2+1}^{n} (n/2)^r \\ &= (n/2) \cdot (n/2)^r \\ &= \Omega(n^{r+1}).\end{aligned}$$

Based on the results, we conclude that the asymptotic tight bound on the summation is

$$\sum_{k=1}^{n} k^r = \Theta(n^{r+1}).$$

b. An asymptotic upper bound:

$$\begin{aligned}\sum_{k=1}^{n} \lg^s k &\leq \sum_{k=1}^{n} \lg^s n \\ &= n \cdot \lg^s n \\ &= O(n \lg^s n).\end{aligned}$$

An asymptotic lower bound:

$$\begin{aligned}\sum_{k=1}^{n} \lg^s k &\geq \sum_{k=n/2+1}^{n} \lg^s k \\ &\geq \sum_{k=n/2+1}^{n} \lg^s (n/2) \\ &= (n/2) \cdot \lg^s (n/2) \\ &= \Omega(n \lg^s n).\end{aligned}$$

The asymptotic tight bound:

$$\sum_{k=1}^{n} \lg^s k = \Theta(n \lg^s n).$$

c. An asymptotic upper bound:

$$\begin{aligned}\sum_{k=1}^{n} k^r \lg^s k &\le \sum_{k=1}^{n} n^r \lg^s n \\ &= n \cdot n^r \lg^s n \\ &= O(n^{r+1} \lg^s n).\end{aligned}$$

An asymptotic lower bound:

$$\begin{aligned}\sum_{k=1}^{n} k^r \lg^s k &\ge \sum_{k=n/2+1}^{n} k^r \lg^s k \\ &\ge \sum_{k=n/2+1}^{n} (n/2)^r \lg^s (n/2) \\ &= (n/2) \cdot (n/2)^r \lg^s (n/2) \\ &= \Omega(n^{r+1} \lg^s n).\end{aligned}$$

The asymptotic tight bound:

$$\sum_{k=1}^{n} k^r \lg^s k = \Theta(n^{r+1} \lg^s n).$$

Assuming $s = 0$ and $r = 0$ in the bound above, respectively, we get the summations and their asymptotic tight bounds from parts (a) and (b).

B Sets, Etc.

B.1 Sets

B.1-1 See Figure B.1-1.

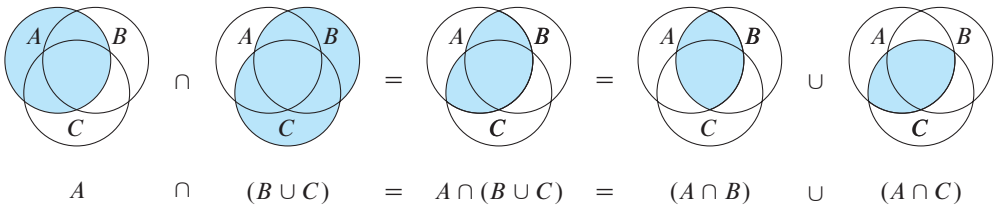

Figure B.1-1 A Venn diagram illustrating the first of the distributive laws (B.1).

B.1-2 We prove the first formula by induction on n. For $n = 1$ the proof is trivial, and for $n = 2$ the formula is the first of DeMorgan's laws. Let's assume that $n > 2$, and that the formula holds for a family of $n - 1$ sets. We have:

$$\begin{aligned}\overline{A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap A_n} &= \overline{(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cap A_n} \\ &= \overline{A_1 \cap A_2 \cap \cdots \cap A_{n-1}} \cup \overline{A_n} \\ &= \overline{A_1} \cup \overline{A_2} \cup \cdots \cup \overline{A_{n-1}} \cup \overline{A_n}.\end{aligned}$$

The second equality follows from the first of DeMorgan's laws, and the third equality follows from the inductive hypothesis.

The proof of the second formula is similar. If $n = 2$, then the formula is DeMorgan's second law, and if $n > 2$, then it is enough to apply the above reasoning with the symbols $\cup$ and $\cap$ swapped, and use DeMorgan's second law.

B.1-3 We prove the principle of inclusion and exclusion by induction on n. For $n = 1$ the proof is trivial, and for $n = 2$ the formula is identical to (B.3). If $n > 2$, then by (B.3) and the generalized first distributive law, we have:

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_n| &= |(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cup A_n| \\ &= |A_1 \cup A_2 \cup \cdots \cup A_{n-1}| + |A_n| \\ &\quad - |(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cap A_n| \\ &= |A_1 \cup A_2 \cup \cdots \cup A_{n-1}| + |A_n| \\ &\quad - |(A_1 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)|.\end{aligned}$$

We now apply the induction hypothesis to the first and the last summand:

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_{n-1}| = &\sum_{1 \le i_1 < n} |A_{i_1}| \\ &- \sum_{1 \le i_1 < i_2 < n} |A_{i_1} \cap A_{i_2}| \\ &+ \sum_{1 \le i_1 < i_2 < i_3 < n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \\ &\qquad \vdots \\ &+ (-1)^{n-2} |A_1 \cap A_2 \cap \cdots \cap A_{n-1}|,\end{aligned}$$

$$\begin{aligned}|(A_1 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)| = &\sum_{1 \le i_1 < n} |A_{i_1} \cap A_n| \\ &- \sum_{1 \le i_1 < i_2 < n} |A_{i_1} \cap A_{i_2} \cap A_n| \\ &\qquad \vdots \\ &+ (-1)^{n-2} |A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap A_n|.\end{aligned}$$

We finish the proof by inserting the resulting expressions into the initial formula:

$$\begin{aligned}|A_1 \cup A_2 \cup \cdots \cup A_n| = &\sum_{1 \le i_1 \le n} |A_{i_1}| \\ &- \sum_{1 \le i_1 < i_2 \le n} |A_{i_1} \cap A_{i_2}| \\ &+ \sum_{1 \le i_1 < i_2 < i_3 \le n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| \\ &\qquad \vdots \\ &+ (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n|.\end{aligned}$$

B.1-4 An example of a one-to-one correspondence between $\mathbb{N}$ and the set of odd natural numbers is: $n \to 2n + 1$. Thus, the set of odd natural numbers is countable.

B.1-5 We prove by induction on the cardinality of the set S.

When $|S| = 0$, it holds that $|2^S| = 2^{|S|} = 1$, because S has exactly one subset — the empty set. Now let $|S| > 0$ and assume that $|2^S| = 2^{|S|}$. Choose $p \notin S$ and consider the set $S' = S \cup \{p\}$. We can divide the subsets of S' into those that contain p and those that do not contain p. By the inductive hypothesis there are $|2^S| = \left|2^{S'-\{p\}}\right| = 2^{|S'-\{p\}|}$ of the latter. In fact there are as many subsets containing p, because each is formed as a union of the singleton $\{p\}$ with some subset not containing p. Thus we have:

$$\begin{aligned}\left|2^{S'}\right| &= 2 \cdot 2^{|S'-\{p\}|} \\ &= 2^{|S'-\{p\}|+1} \\ &= 2^{|S'|}.\end{aligned}$$

B.1-6 An n-tuple of elements $a_1, a_2, \ldots, a_n$ is defined as:

$$(a_1, a_2, \ldots, a_n) = \begin{cases} \emptyset & \text{if } n = 0, \\ \{a_1\} & \text{if } n = 1, \\ \{a_1, \{a_1, a_2\}\} & \text{if } n = 2, \\ (a_1, (a_2, \ldots, a_n)) & \text{if } n \geq 3. \end{cases}$$

B.2 Relations

B.2-1 To show that the relation "$\subseteq$" on the power set $2^{\mathbb{Z}}$ is a partial order, we will show that "$\subseteq$" is reflexive, antisymmetric and transitive. Let $A, B, C \in 2^{\mathbb{Z}}$. Of course, $x \in A$ implies $x \in A$, so reflexivity trivially holds. If $A \subseteq B$ and $B \subseteq A$, then $x \in A$ implies $x \in B$, and $x \in B$ implies $x \in A$. By the law of transitivity of implication, $x \in A$ if and only if $x \in B$, and therefore $A = B$. The combination of $A \subseteq B$ and $B \subseteq C$ means that $x \in A$ implies $x \in B$, and $x \in B$ implies $x \in C$. Referring again to the law of transitivity of implication, we have that $x \in A$ implies $x \in C$, that is, $A \subseteq C$.

The relation "$\subseteq$" on $2^{\mathbb{Z}}$ is not a total order, because, for example, $\{0, 1\} \not\subseteq \{1, 2\}$ and $\{1, 2\} \not\subseteq \{0, 1\}$.

B.2-2 Let us denote the relation "equivalent modulo n", for $n \in \mathbb{N} - \{0\}$, by

$$R_n = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : a = b \pmod{n}\}.$$

For any $a \in \mathbb{Z}$ we have $a = a \pmod{n}$, because $a - a = 0 \cdot n$, so the relation R_n is reflexive. For any $a, b \in \mathbb{Z}$, if there exists $q \in \mathbb{Z}$ that $a - b = qn$, then $b - a = -qn$, and therefore the fact that $a = b \pmod{n}$ implies $b = a \pmod{n}$. Therefore, R_n is symmetric. For transitivity let's choose any $a, b, c \in \mathbb{Z}$ such that $a = b \pmod{n}$ and $b = c \pmod{n}$. This means that there exist $q, r \in \mathbb{Z}$ such that $a - b = qn$ and $b - c = rn$. Hence $a - c = a - b + b - c = qn + rn = (q + r)n$, and therefore $a = c \pmod{n}$.

Based on the proof above R_n is an equivalence relation that partitions the set $\mathbb{Z}$ into n abstraction classes; the ith class, where $i = 1, 2, \ldots, n$, is the set of integers that leave remainder $i - 1$ when divided by n.

B.2-3

a. The relation $\{(a, a), (b, b), (c, c), (a, b), (b, a), (a, c), (c, a)\}$ on the set $\{a, b, c\}$.

b. The relation $\{(a, a), (a, b), (b, b)\}$ on the set $\{a, b\}$.

c. The empty relation on any nonempty set.

B.2-4 If R is an equivalence relation, then $s \in [s]$ for all $s \in S$. By antisymmetry of R, if $s' \, R \, s$ and $s \, R \, s'$, then $s = s'$, so there are no such s' that $s' \in [s] - \{s\}$. This means that for all $s \in S$, $[s] = \{s\}$.

B.2-5 Symmetry and transitivity are defined using implications. For an implication "p implies q" to hold true, it isn't required for p to be true. A relation on a set remains symmetric and transitive, if there is an element in the set, that is not related to any element in that set. Reflexivity, on the other hand, requires that each element is related to itself. Thus, there exist relations that are symmetric and transitive relations but not reflexive (an example of one is given in Exercise B.2-3(c)).

B.3 Functions

B.3-1

a. For any function $f : A \to B$, $f(A) \subseteq B$, which results directly from the definition of the range of f. Additionally, if f is injective, then each argument $a \in A$ has a distinct image $f(a) \in f(A)$, so $|A| = |f(A)|$. Hence, $|A| \leq |B|$.

b. For any function $f : A \to B$, where A is finite, it holds $|A| \geq |f(A)|$, because there may be arguments $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$. Additionally, if f is surjective, then $f(A) = B$, and therefore $|A| \geq |B|$.

B.3-2 The function $f(x) = x + 1$, with $\mathbb{N}$ as both the domain and the codomain, is not bijective because there is no $x \in \mathbb{N}$ such that $f(x) = 0$. If we consider $\mathbb{Z}$ as the domain and the codomain, then f becomes bijective, since every integer y in the codomain is an image under f of a uniquely determined integer in the domain: $y = f(y - 1)$.

B.3-3 The definition comes as a natural generalization of the functional inverse. Let R be a binary relation on sets A an B. The binary relation R^{-1} on sets B and A, such that

$b\ R^{-1}\ a$ if and only if $a\ R\ b$,

is called the ***inverse*** of R.

B.3-4 Let $h : \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ be the function we are looking for. Since the inverse of a bijection is also a bijection, we will focus on finding the inverse $h^{-1} : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, which is equivalent to finding a way to sequentially number each ordered pair with integer elements, so that each integer is used as the number of some pair. We will now describe the construction of one of such mappings.

Let us make some simplification — instead of numbering pairs by integers, we will restrict ourselves to natural numbers. Let $g : \mathbb{Z} \times \mathbb{Z} \to \mathbb{N}$ and $f : \mathbb{N} \to \mathbb{Z}$ be such bijections that $h^{-1} = f \circ g$. It was already noted on page 1162 of the text that $f(n) = (-1)^n \lceil n/2 \rceil$ is bijective, so all that remains is to find the function g.

Consider the numbering of ordered pairs with integer elements shown in Figure B.3-4 in the form of a spiral. Since each such pair (x, y) is assigned a unique natural number, we can treat this spiral as a description of the function g. Let $d = \max\{|x|, |y|\}$ and $D = (2d - 1)^2 - 1$. Informally, d and D denote, respectively, the number of the "lap" around the origin covered by the spiral at the time of passing through (x, y), and the largest number on the spiral during the previous "lap". Then we can define the function g as follows:

$$g(x, y) = \begin{cases} 0 & \text{if } d = |x| = |y| = 0, \\ D + \ d + y & \text{if } d = x \neq |y|, \\ D + 3d - x & \text{if } d = y \neq 0, \\ D + 5d - y & \text{if } d = -x \neq |y|, \\ D + 7d + x & \text{if } d = -y \neq 0. \end{cases}$$

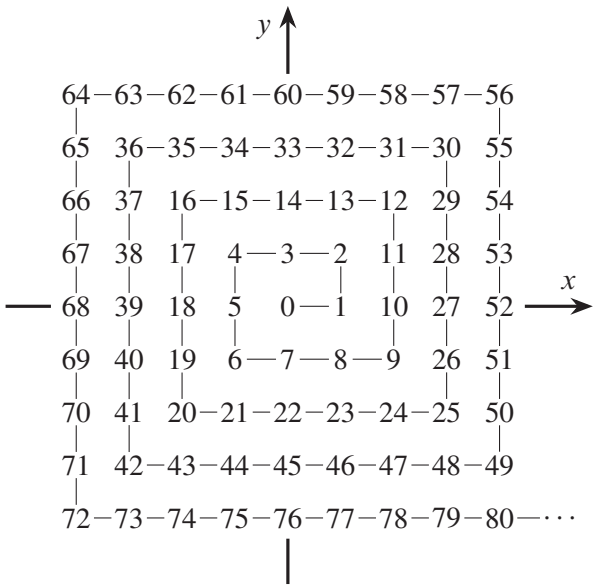

Figure B.3-4 A bijection from $\mathbb{Z}\times\mathbb{Z}$ to $\mathbb{N}$. Individual natural numbers denote the values of this bijection for points with integer coordinates in the Cartesian coordinate system.

B.4 Graphs

B.4-1 We can represent the relation of "shaking hands" on the faculty party as an undirected graph $G = (V, E)$, where V is the set of professors attending the party and E is the set of unordered pairs of professors who shook hands. Summing up the degrees of all vertices of G, we get twice the number of its edges, because each edge is incident on exactly two vertices.

B.4-2 Let $\langle v_0, v_1, \ldots, v_n\rangle$ be a path from vertex u to vertex v in a directed or undirected graph. If the path is simple, the vertices in the path are distinct. Otherwise, the path contains a cycle $\langle v_i, v_{i+1}, \ldots, v_{i+k}, v_i\rangle$ for some i and k, and by eliminating the subpath $\langle v_i, v_{i+1}, \ldots, v_{i+k}\rangle$ from the path, we discard an extra repetition of v_i. By applying this modification until we remove all such repetitions, we obtain a simple path from u to v.

The proof for cycles is analogous with $v_0 = v_n$ and we have to remember not to remove the very last repetition of v_0, required for the path to form a cycle.

B.4-3 We proceed by induction on the number of vertices in the graph $G = (V, E)$. When $|V| = 1$, the graph has no edges and the inequality trivially holds. For the inductive

step, pick a vertex $v \in V$. G is connected so there must be at least one edge incident on v and another vertex from V. Let $G' = (V', E')$ be the subgraph of G induced by $V' = V - \{v\}$ (i.e., the graph G with vertex v and all edges incident on v removed). We have $|V'| = |V| - 1$ and $|E'| \leq |E| - 1$, so:

$$\begin{aligned} |E| &\geq |E'| + 1 \\ &\geq (|V'| - 1) + 1 \qquad \text{(by the inductive hypothesis applied for } G') \\ &= |V| - 1. \end{aligned}$$

B.4-4 Every vertex of a directed or undirected graph is reachable from itself because there is a path of length 1 containing only that vertex, therefore the "is reachable from" relation is reflexive.

Let u, v, w be vertices of a directed or undirected graph such that $u \overset{p}{\rightsquigarrow} v$ and $v \overset{q}{\rightsquigarrow} w$. Then there is a path from u to w which is a concatenation of the sequences p and q (with the extra v between them eliminated), which proves transitivity.

The relation is symmetric only in an undirected graph, because for its any vertices u, v, if $u \overset{p}{\rightsquigarrow} v$ then we can mirror the sequence p to obtain p' such that $v \overset{p'}{\rightsquigarrow} u$. The sequence p' is a valid path, because if (x, y) is an edge, so is (y, x). In a directed graph the edges are ordered pairs, so if (x, y) is an edge, (y, x) may not be. Thus mirroring may not produce a valid path and so symmetry does not hold in general for directed graphs.

B.4-5 See Figure B.4-5.

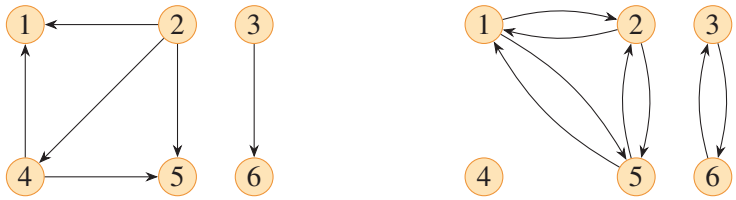

Figure B.4-5 **(a)** The undirected version of the directed graph in Figure B.2(a). **(b)** The directed version of the undirected graph in Figure B.2(b).

B.4-6 A hypergraph $H = (V_H, E_H)$ can be represented as a bipartite graph $G = (V_1 \cup V_2, E)$, where $V_1 = V_H$ and $V_2 = E_H$. A pair $\{v_1, v_2\}$, such that $v_1 \in V_1$ and $v_2 \in V_2$, is an edge of the graph G, if and only if the hyperedge v_2 is incident on vertex v_1 in the hypergraph H. In the graph G there are no edges between elements of V_1 or between elements of V_2, so G is indeed bipartite.

B.5 Trees

B.5-1 See Figure B.5-1.

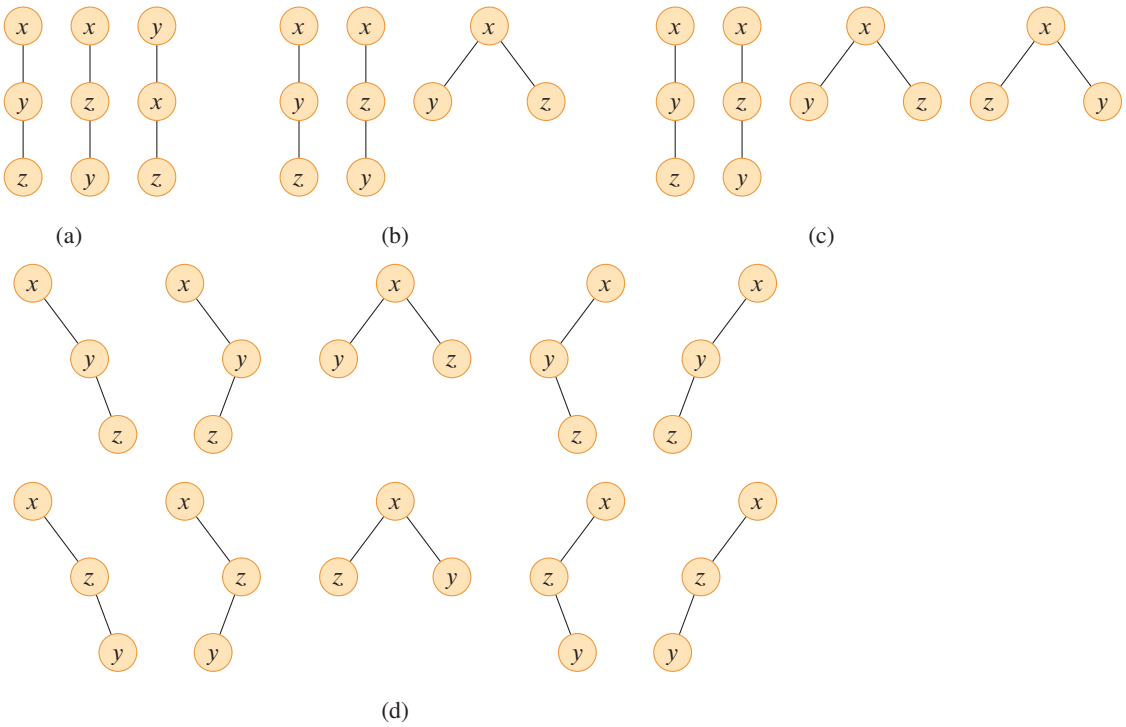

Figure B.5-1 **(a)** All the free trees with vertices x, y, and z. **(b)** All the rooted trees with nodes x, y, and z with x as the root. **(c)** All the ordered trees with nodes x, y, and z with x as the root. **(d)** All the binary trees with nodes x, y, and z with x as the root.

B.5-2 Let $G' = (V, E')$ be the undirected version of the graph G. Note that G' is connected, since every vertex $v \in V$ is reachable from vertex v_0 in G'.

Now suppose that G' does not form a tree. This means it has a cycle of length $k \geq 3$, in particular a simple cycle of such a length (after Exercise B.4-2), say $\langle v_1, v_2, \ldots, v_{k+1} \rangle$, where $v_{k+1} = v_1$. For simplicity, let $v_{k+2} = v_2$. The graph G is acyclic, so for some integer $1 \leq l \leq k$ there must be edges (v_l, v_{l+1}), $(v_{l+2}, v_{l+1}) \in E$. By assumption, both v_l and v_{l+2} are reachable from v_0, so there are two different paths from v_0 to v_{l+1}: $v_0 \rightsquigarrow v_l \rightarrow v_{l+1}$ and $v_0 \rightsquigarrow v_{l+2} \rightarrow v_{l+1}$. The obtained contradiction leads to the conclusion that G' is acyclic and so it is a tree.

B.5-3 The proof is by induction on the number n of nodes in the tree. The base case is when $n = 1$. A binary tree with only one node has one leaf and no degree-2 nodes, thus the statement holds.

For the inductive step, suppose that $n \geq 1$ and assume that the statement is true for all binary trees with n nodes. Let T be a binary tree with $n + 1$ nodes. We can remove the root node and obtain two subtrees, each of which has less than $n + 1$ nodes. If any of the subtrees is empty, the statement follows immediately from the inductive hypothesis applied to the other (nonempty) subtree, since the root of T is neither a leaf nor a degree-2 node.

Now suppose that both the left subtree and the right subtree are nonempty. Let L and D be, respectively, the number of leaves and the number of degree-2 nodes in T. Similarly, let L_1 and D_1 be the analogous numbers for the left subtree, and let L_2 and D_2 be the analogous numbers for the right subtree. It holds that $L = L_1 + L_2$ and $D = D_1 + D_2 + 1$. Then:

$$\begin{aligned} D &= D_1 + D_2 + 1 \\ &= L_1 - 1 + L_2 - 1 + 1 && \text{(by the inductive hypothesis)} \\ &= L_1 + L_2 - 1 \\ &= L - 1. \end{aligned}$$

In a full binary tree each internal node is a degree-2 node, so based on the above result, such a tree has one fewer internal nodes than leaves.

B.5-4 We construct a sequence of full binary trees $T_1, T_2, \ldots$, as follows. Let T_1 be a binary tree containing a single node, which obviously is a single leaf. For any $k \geq 2$, let T_k be a binary tree whose left subtree is T_{k-1} and whose right subtree is a one-node binary tree. Given that T_{k-1} has $k - 1$ leaves, T_k has $(k - 1) + 1 = k$ leaves.

B.5-5 The proof is by induction on the height h of the binary tree.

The only nonempty binary tree with $h = 0$ is the one with a single node, so $n = 1$ and the statement $h \geq \lfloor \lg n \rfloor$ holds. Now let $h > 0$ and suppose that the statement holds for all binary trees with heights at most $h - 1$. Consider a binary tree T with height h. One of its subtrees must have height $h' = h - 1$, and the other subtree must have height $h'' \leq h - 1$. Let n' and n'' be the numbers of nodes in these subtrees, respectively. By the inductive hypothesis, $h' \geq \lfloor \lg n' \rfloor$ and $h'' \geq \lfloor \lg n'' \rfloor$. Observe that for any integer $k \geq 0$ it is true that $\lfloor \lg 2^{k+1} \rfloor = k + 1$ but $\lfloor \lg(2^{k+1} - 1) \rfloor = k$. Therefore we have

$$n' \leq 2^{h'+1} - 1$$

and

$$n'' \le 2^{h''+1} - 1.$$

The tree T has $n = n' + n'' + 1$ nodes, so

$$\begin{aligned} n &= n' + n'' + 1 \\ &\le 2^{h'+1} - 1 + 2^{h''+1} - 1 + 1 \\ &\le 2^h + 2^h - 1 \\ &= 2^{h+1} - 1. \end{aligned}$$

Using the above observation again, we conclude that $\lfloor \lg n \rfloor \le h$.

B.5-6 We proceed by induction on n. The base case is when $n = 0$. The only full binary tree that has no internal nodes is a one-node tree, for which we have $e = i = 0$ and the statement holds.

Now let $n \ge 1$ and suppose that the statement holds for all full binary trees with less than n internal nodes. We can construct an arbitrary full binary tree T with n internal nodes by choosing any pair of full binary trees, whose numbers of internal nodes sum up to $n - 1$, for its left subtree T_L and its right subtree T_R. For $k \ge 0$, if T_L has k internal nodes, T_R has $n - k - 1$ internal nodes. By the conclusion made in Exercise B.5-3, T_L and T_R have $k + 1$ and $n - k$ leaves, respectively.

Let i_L and i_R be the internal path lengths of T_L and T_R, respectively, and similarly, let e_L and e_R be their external path lengths. The depth of each node in T — other than the root, for which the depth is 0 — is 1 more than the depth of this node in either T_L or T_R. Hence,

$$\begin{aligned} i &= 0 + (i_L + k) + (i_R + n - k - 1) \\ &= i_L + i_R + n - 1 \end{aligned}$$

and

$$\begin{aligned} e &= (e_L + k + 1) + (e_R + n - k) \\ &= e_L + e_R + n + 1. \end{aligned}$$

By the inductive hypothesis, $e_L = i_L + 2k$ and $e_R = i_R + 2(n - k - 1)$, so

$$\begin{aligned} e - i &= e_L + e_R + n + 1 - i_L - i_R - n + 1 \\ &= i_L + 2k + i_R + 2(n - k - 1) - i_L - i_R + 2 \\ &= 2n. \end{aligned}$$

Therefore, the statement holds for all full binary trees.

B.5-7 The proof is by induction on the number n of nodes of the tree. In the empty tree (where $n = 0$) the sum of the leaf weights is 0, which obviously satisfies the inequality. When $n = 1$, then the tree consists of a single leaf with a depth of 0, so the sum of the leaf weights is $2^{-0} = 1$.

Now suppose that $n > 1$ and that the inequality holds for trees with less than n nodes. Let T be a binary tree with n nodes, and let T_L, T_R be its left and right subtrees, respectively. By the inductive hypothesis, we have that the inequality is satisfied separately for trees T_L and T_R. The depth of a leaf in tree T is 1 greater than the depth of the same leaf in either T_L or T_R, hence the weights of the leaves of T are half as compared to the weights of the same leaves in its subtrees. Thus, in tree T both the sum of the leaf weights from T_L, as well as the sum of the leaf weights from T_R do not exceed $1/2$, and so the total sum of the leaf weights in T does not exceed 1.

B.5-8 Let T be a binary tree with $L \geq 2$ leaves. Suppose that T does not contain a subtree having between $L/3$ and $2L/3$ leaves, inclusive. Since $L \geq 2$, the root of T is not a leaf. We will follow a path p from the root to a leaf, at each level traversing from a node to its child whose subtree has more leaves. The numbers of leaves in subtrees rooted at the visited nodes decrease from L to 1 as we move on the path p. By assumption, there is a node x on path p such that the subtree rooted at x has more than $2L/3$ leaves, and that the subtree T_y rooted at its child y has less than $L/3$ leaves. The subtree rooted at the other child of x has fewer leaves than T_y (0 if x has only one child). This leads to a contradiction because the total number of leaves in both subtrees is less than $2L/3$.

Problems

B-1 ***Graph coloring***

a. For a given tree let us distinguish any of its nodes and view the tree as a rooted tree in that node. Edges in such a tree are incident on nodes from adjacent levels, so nodes can be colored according to the parity of their depth in the tree (e.g., those on odd levels get color 1, and those on even levels get color 2).

b. (1) $\Rightarrow$ (2): Since $G = (V, E)$ is a bipartite graph, the vertex set V can be partitioned into two subsets V_1 and V_2, such that vertices from one subset aren't adjacent to vertices from the other subset. Thanks to this property it is possible to assign one color to each vertex from V_1 and a different color to each vertex from V_2, obtaining a 2-coloring of graph G.

(2) $\Rightarrow$ (3): Suppose that graph G has a cycle $p = \langle v_0, v_1, \ldots, v_{2k+1} \rangle$ for some $k \geq 1$. Let c be a 2-coloring of G and without loss of generality suppose that

$c(v_0) = 1$. Then it must be $c(v_{2i}) = 1$ and $c(v_{2i+1}) = 2$ for all $i = 0, 1, \ldots, k$. But since p is a cycle, $v_{2k+1} = v_0$, which yields the contradiction. Therefore, graph G must not contain a cycle of odd length.

(3) $\Rightarrow$ (1): Suppose that graph G is connected. Otherwise, apply the following proof separately to each connected component of G.

Choose any $v_0 \in V$. Let V_1 be the set of all vertices $v \in V$ for which there exists a path from v_0 to v of even length, and let $V_2 = V - V_1$. If there were vertices u, $w \in V_1$, such that there is a path from u to w of odd length, then $v_0 \rightsquigarrow u \rightsquigarrow w \rightsquigarrow v_0$ would be a cycle of an odd length. Therefore, no two vertices from V_1 are adjacent, and we can prove a similar fact for vertices in V_2. This means that $G = (V_1 \cup V_2, E)$ is bipartite.

c. We carry out the proof by induction on the number of vertices in graph $G = (V, E)$. If the graph has only one vertex (with degree 0), then of course one color is enough.

Now suppose that $|V| > 1$. Let us choose an arbitrary vertex $v \in V$ and denote by G' the subgraph of G induced by the set $V - \{v\}$. By the inductive hypothesis, G' can be colored with $d' + 1$ colors, where d' is the maximum degree of any vertex in G'. It is true that $d' \leq d$, so G' can also be colored with $d + 1$ colors. Since vertex v has at most d neighbors in G, among the colors used in a $(d + 1)$-coloring of G' there is one that is not assigned to any neighbor of v in graph G. By assigning that color to v, we obtain a $(d + 1)$-coloring of G.

d. For any $k \geq 2$, if graph G is k-colorable, but is not $(k - 1)$-colorable, then for every two different colors used in a k-coloring of G there are two adjacent vertices, one with the first color and the other with the second color. If that weren't true, there would be two different colors that we could not distinguish, and so we could reduce the number of colors required, obtaining a $(k - 1)$-coloring of G.

Based on the above reasoning, we have $\binom{k}{2} \leq |E|$. We use the fact that for $k \geq 2$ it holds that $k/2 \leq k - 1$, and hence

$$\begin{aligned} k^2 &= 4 \cdot \frac{k}{2} \cdot \frac{k}{2} \\ &\leq 4 \cdot \frac{k(k-1)}{2} \\ &= 4\binom{k}{2} \\ &\leq 4|E| \\ &= O(|V|). \end{aligned}$$

Therefore, $k = O(\sqrt{|V|})$.

B-2 *Friendly graphs*

a.

Theorem
In any undirected graph $G = (V, E)$, where $|V| \geq 2$, there are two vertices with the same degree.

Proof The degrees of the vertices in graph G are numbers from 0 to $|V|-1$, inclusive. Observe, however, that G has an isolated vertex if and only if it doesn't have a vertex adjacent to any other vertex of G. Therefore, the degrees of the vertices of G can in fact take $|V| - 1$ possible values, so some two vertices must have equal degrees. ■

b.

Theorem
An undirected graph $G = (V, E)$ of 6 vertices, or its complement[1] $\overline{G}$, contain a clique[2] of size 3.

Proof Let's pick any $v \in V$. Since $|V| = 6$, there is a set $U \subseteq V - \{v\}$ of three vertices, for which exactly one of the following cases holds:

1. each vertex in U is adjacent to v,
2. no vertex in U is adjacent to v.

Since case 1 in G is equivalent to case 2 in $\overline{G}$, and vice versa, let us focus on case 1 only. If there is a pair of adjacent vertices among those in U, then the vertices of this pair together with v form a clique in G. Now suppose there is no such pair. Then, however, the vertices in U form a clique in $\overline{G}$. ■

c.

Theorem
For any undirected graph $G = (V, E)$ the set V can be partitioned into two subsets such that for any vertex $v \in V$ at least half of its neighbors do not belong to the subset that v belongs to.

---

[1] See page 1085 of the text.
[2] See page 1081 of the text.

Proof Let us divide the vertices of G arbitrarily into two disjoint subsets V_1 and V_2. Let

$$S = \{(v_1, v_2) \in E : v_1 \in V_1,\ v_2 \in V_2\}.$$

If more than half of the neighbors of some vertex $v \in V_1$ are also in V_1, then moving v to V_2 increases the size of set S. Similarly, for vertices $v \in V_2$ with this property — by moving them to set V_1, we increase set S. This process must terminate after a finite number of steps, because S cannot grow indefinitely. The obtained partition $V = V_1 \cup V_2$ satisfies the desired property. ■

d. Before we formulate the main theorem, we will need the following lemma.

Lemma
Let $G = (V, E)$ be an undirected graph, where $|V| \geq 3$ and

$$\text{degree}(u) + \text{degree}(v) \geq |V| \tag{1}$$

for every pair of non-adjacent vertices $u, v \in V$. Then G is hamiltonian[3].

Proof Suppose the lemma does not hold, that is there is a graph $G = (V, E)$ with at least 3 vertices, that satisfies the property (1), but is not hamiltonian. Among all such graphs let us consider that of the greatest number of edges $|E|$. Let $n = |V|$. Then G must have a hamiltonian path[4] $\langle v_1, v_2, \ldots, v_n \rangle$ — otherwise we could add the missing edges without violating (1) and get a graph with more than $|E|$ edges. Since G does not have a hamiltonian cycle, $(v_1, v_n) \notin E$. By assumption, we have that $\text{degree}(v_1) + \text{degree}(v_n) \geq n$.

Now let us define the following sets:

$$S_1 = \{i : 2 \leq i \leq n \text{ and } (v_1, v_i) \in E\},$$
$$S_n = \{i : 2 \leq i \leq n \text{ and } (v_{i-1}, v_n) \in E\}.$$

Then $|S_1| = \text{degree}(v_1)$ and $|S_n| = \text{degree}(v_n)$. Since $|S_1| + |S_n| \geq n$ and the set $S_1 \cup S_n$ has at most $n-1$ elements, the set $S_1 \cap S_n$ must be non-empty. So there exists i for which $(v_1, v_i), (v_{i-1}, v_n) \in E$. Hence, the path

$$v_1 \rightsquigarrow v_{i-1} \to v_n \to v_{n-1} \rightsquigarrow v_i \to v_1$$

is a hamiltonian cycle in graph G, which is a contradiction. ■

Theorem
Let $G = (V, E)$ be an undirected graph, where $|V| \geq 3$ and

$$\text{degree}(v) \geq |V|/2$$

[3] See page 1056 of the text.
[4] See Exercise 34.2-6.

for every vertex $v \in V$. Then G is hamiltonian.

Proof If the property holds for every $v \in V$, then

$$\text{degree}(u) + \text{degree}(v) \geq |V|$$

for every $u, v \in V$ whether they are adjacent or not. So G satisfies the assumptions of the lemma, and thus it is hamiltonian. ■

B-3 *Bisecting trees*

a. The fact trivially holds for binary trees with at most three nodes, so in the rest of the proof we will assume that $n \geq 4$. We define a ***branch*** of a binary tree as a path $\langle x_0, x_1, \ldots, x_k \rangle$, where x_0 is the root, x_k is a leaf, and for $i = 1, 2, \ldots, k$, x_i is the root of the subtree of x_{i-1} with more nodes. Note that a tree may have more than one branch. Let's denote by $s(x)$ the number of nodes in the subtree rooted at x. For every $i = 0, 1, \ldots, k-1$,

$$s(x_i) \leq 2s(x_{i+1}) + 1. \tag{2}$$

Of course, the sequence $\langle s(x_0), s(x_1), \ldots, s(x_k) \rangle$ decreases from n to 1, so there must be j, such that $s(x_j) > n/4$ and $s(x_{j+1}) \leq n/4$. Thus, by removing edge (x_{j-1}, x_j), we partition the nodes of the tree into two sets of sizes

$$\begin{aligned} s(x_j) &\leq 2s(x_{j+1}) + 1 && \text{(by inequality (2))} \\ &= 2n/4 + 1 \\ &\leq 3n/4 \end{aligned}$$

and

$$\begin{aligned} n - s(x_j) &< n - n/4 \\ &= 3n/4. \end{aligned}$$

b. The constant $3/4$ is sufficient to make balanced partitions, as we have shown in part (a). The example of the binary tree in Figure B-3 shows that the constant cannot be lower.

c. Assume that $n \geq 6$. We'll remove one edge from the original tree to cut off a subtree of at most $\lfloor n/2 \rfloor$ vertices. Consider any of the tree's branches $\langle x_0, x_1, \ldots, x_k \rangle$, as well as the function s, both defined in the solution to part (a). There is an edge (x_j, x_{j+1}), such that $s(x_j) > \lfloor n/2 \rfloor$ and $s(x_{j+1}) \leq \lfloor n/2 \rfloor$. From inequality (2) we

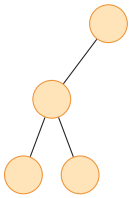

Figure B-3 The binary tree whose most evenly balanced partition upon removal of a single edge has a subset with 3 vertices.

have

$$\begin{aligned} s(x_{j+1}) &\geq (s(x_j) - 1)/2 \\ &> (\lfloor n/2 \rfloor - 1)/2 \\ &\geq \lfloor n/2 \rfloor / 3. \end{aligned}$$

We cut off the subtree rooted at x_{j+1} and continue partitioning the remaining tree, with at most

$$\lfloor n/2 \rfloor - s(x_{j+1}) < (2/3)\lfloor n/2 \rfloor$$

vertices still to remove from it. In each subsequent step the number of nodes remaining to be cut is reduced by a factor of $2/3$. Therefore, we'll need to remove $O(\log_{3/2} n) = O(\lg n)$ edges.

C Counting and Probability

C.1 Counting

C.1-1 We will only count non-empty k-substrings, so $k \geq 1$. The first k-substring occupies positions 1 through k in the n-string, the second one occupies positions 2 through $k+1$, and so on. The last k-substring ends at position n, so it must start at position $n-k+1$. So an n-string has $n-k+1$ k-substrings in total.

By the rule of sum, the total number of all substrings of an n-string is

$$\begin{aligned}\sum_{k=1}^{n}(n-k+1) &= \sum_{k=1}^{n} k \\ &= \frac{n(n+1)}{2} && \text{(by equation (A.1))} \\ &= \binom{n+1}{2}.\end{aligned}$$

C.1-2 Consider the set of n-bit integers $N = \{0, 1, \ldots, 2^n - 1\}$ and the set of m-bit integers $M = \{0, 1, \ldots, 2^m - 1\}$. We are counting functions $f : N \to M$ or, equivalently, sequences $\langle a_1, \ldots, a_{2^n} \rangle$ with terms from M. We can choose each term in such a sequence in 2^m ways, so following the rule of product gives us

$$(2^m)^{2^n} = 2^{m2^n}$$

possibilities in total for the number of sequences or the number of n-input, m-output boolean functions. Specifically, there is

$$2^{2^n}$$

n-input, 1-output boolean functions.

C.1-3 Let S_n be the number of ways to arrange n professors around a circular table. A seating is indistinguishable from $n-1$ others that are formed from it by rotation. There are $n!$ permutations of the n professors, so $nS_n = n!$. Hence

$$S_n = (n-1)!.$$

C.1-4 Three integers add up to an even number only if all of them are even, or if exactly one of them is even. There are 49 even and 50 odd numbers in the set $\{1, 2, \ldots, 99\}$, so in the first option we can make a choice in $\binom{49}{3}$ ways, and in the second — in $\binom{50}{2}\binom{49}{1}$ ways. Therefore, the total number of ways is

$$\binom{49}{3} + \binom{50}{2}\binom{49}{1} = 78{,}449.$$

C.1-5

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!\,(n-k)!} \\ &= \frac{n}{k} \cdot \frac{(n-1)!}{(k-1)!\,(n-k)!} \\ &= \frac{n}{k}\binom{n-1}{k-1}\end{aligned}$$

C.1-6

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!\,(n-k)!} \\ &= \frac{n}{n-k} \cdot \frac{(n-1)!}{k!\,(n-k-1)!} \\ &= \frac{n}{n-k}\binom{n-1}{k}\end{aligned}$$

C.1-7 Let S be an n-set and let $s \in S$. The set S has $\binom{n}{k}$ k-subsets, for $0 \le k \le n$. We can divide these k-subsets into those that do not contain s and those that do contain s. There are $\binom{n-1}{k}$ of the former and $\binom{n-1}{k-1}$ of the latter. Hence, we have

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

C.1-8 The first rows of Pascal's triangle:

$$
\begin{array}{ccccccc}
1 & & & & & & \\
1 & 1 & & & & & \\
1 & 2 & 1 & & & & \\
1 & 3 & 3 & 1 & & & \\
1 & 4 & 6 & 4 & 1 & & \\
1 & 5 & 10 & 10 & 5 & 1 & \\
1 & 6 & 15 & 20 & 15 & 6 & 1
\end{array}
$$

In the nth row the leftmost element is $\binom{n}{0} = 1$, and the rightmost element is $\binom{n}{n} = 1$. The remaining elements are determined based on the identity from Exercise C.1-7.

C.1-9

$$
\begin{aligned}
\binom{n+1}{2} &= \frac{(n+1)!}{2!\,(n-1)!} \\
&= \frac{n(n+1)}{2} \\
&= \sum_{i=1}^{n} i \qquad \text{(by equation (A.1))}
\end{aligned}
$$

C.1-10 Let's fix n. We'll view the expression $\binom{n}{k}$ as a function $b_n(k)$ for $k = 0, 1, \ldots, n$ and will find k for which the value $b_n(k)$ is the greatest. Consider the ratio:

$$
\begin{aligned}
\frac{b_n(k+1)}{b_n(k)} &= \frac{\binom{n}{k+1}}{\binom{n}{k}} \\
&= \frac{n!}{(k+1)!\,(n-k-1)!} \cdot \frac{k!\,(n-k)!}{n!} \\
&= \frac{n-k}{k+1}.
\end{aligned}
$$

If $n - k \geq k + 1$, or $k \leq (n-1)/2$, then function b_n is monotonically increasing. Conversely, when $k \geq (n-1)/2$, then function b_n is monotonically decreasing. So

when n is odd, b_n reaches its maximum for $k = (n-1)/2 = \lfloor n/2 \rfloor$. Moreover:

$$\begin{aligned} b_n((n-1)/2) &= \binom{n}{(n-1)/2} \\ &= \binom{n}{n-(n-1)/2} && \text{(by equation (C.3))} \\ &= \binom{n}{(n+1)/2} \\ &= b_n((n+1)/2), \end{aligned}$$

so the maximum value is also reached when $k = (n+1)/2 = \lceil n/2 \rceil$.

In case n is an even number, b_n achieves the maximum value for either $k = n/2$ or $k = n/2 - 1$. Let's see which value is bigger:

$$\begin{aligned} \frac{b_n(n/2)}{b_n(n/2-1)} &= \frac{\binom{n}{n/2}}{\binom{n}{n/2-1}} \\ &= \frac{n!}{(n/2)!\,(n/2)!} \cdot \frac{(n/2-1)!\,(n/2+1)!}{n!} \\ &= \frac{n/2+1}{n/2} \\ &> 1. \end{aligned}$$

So we have that b_n reaches its maximum for $k = n/2 = \lfloor n/2 \rfloor = \lceil n/2 \rceil$.

C.1-11 First, let's observe that

$$\begin{aligned} (j+k)! &= j!\,(j+1)(j+2)\cdots k \\ &\geq j!\,1 \cdot 2 \cdots k \\ &= j!\,k!, \end{aligned}$$

where equality holds only if $j = 0$ or $k = 0$. Then

$$\begin{aligned}\binom{n}{j+k} &= \frac{n!}{(j+k)!\,(n-j-k)!} \\ &\le \frac{n!}{j!\,k!\,(n-j-k)!} \\ &= \frac{n!}{j!\,(n-j)!} \cdot \frac{(n-j)!}{k!\,(n-j-k)!} \\ &= \binom{n}{j}\binom{n-j}{k}.\end{aligned}$$

We can interpret the expression $\binom{n}{j+k}$ as the number of possible ways to choose $j + k$ items out of n, and the expression $\binom{n}{j}\binom{n-j}{k}$ as the number of ways to choose j items out of n, and then k items out of $n - j$ left after the first choice. In each strategy we come up with a set of the $j + k$ selected items. There is exactly one way to construct a given set of $j + k$ items using the first strategy, and at least one if we follow the second strategy. It is so, because we can arbitrarily partition the set into the set of j items that will be selected in the first step and the set of k items that will be selected in the second step.

C.1-12 It's easy to check that the inequality holds for $k = 0$. Now let $k \ge 1$ and let's assume that

$$\binom{n}{k-1} \le \frac{n^n}{(k-1)^{k-1}(n-k+1)^{n-k+1}}.$$

Then

$$\begin{aligned}\binom{n}{k} &= \frac{n}{k}\binom{n-1}{k-1} && \text{(by Exercise C.1-5)} \\ &= \frac{n}{k} \cdot \frac{n-k+1}{n}\binom{n}{k-1} && \text{(by Exercise C.1-6)} \\ &\le \frac{n^n}{k(k-1)^{k-1}(n-k+1)^{n-k}} && \text{(by the inductive hypothesis).}\end{aligned}$$

Now it suffices to show that

$$\frac{n^n}{k(k-1)^{k-1}(n-k+1)^{n-k}} \le \frac{n^n}{k^k(n-k)^{n-k}}.$$

To prove this, we examine the ratio of both expressions:

$$\begin{aligned}\frac{\frac{n^n}{k(k-1)^{k-1}(n-k+1)^{n-k}}}{\frac{n^n}{k^k(n-k)^{n-k}}} &= \frac{k^{k-1}(n-k)^{n-k}}{(k-1)^{k-1}(n-k+1)^{n-k}} \\ &= \frac{\left(\frac{k}{k-1}\right)^{k-1}}{\left(\frac{n-k+1}{n-k}\right)^{n-k}} \\ &= \frac{\left(1+\frac{1}{k-1}\right)^{k-1}}{\left(1+\frac{1}{n-k}\right)^{n-k}}.\end{aligned}$$

The sequence $e_n = (1+1/n)^n$ is increasing — therefore the above ratio does not exceed 1, as long as $k-1 \le n-k$ or, equivalently, $k \le (n+1)/2$, and in particular when $k \le n/2$. Thus, we have shown that inequality (C.7) holds for $k \le n/2$.

When $n/2 \le k \le n$, then $0 \le n-k \le n/2$, so

$$\begin{aligned}\binom{n}{k} &= \binom{n}{n-k} && \text{(by equation (C.3))} \\ &\le \frac{n^n}{(n-k)^{n-k}k^k},\end{aligned}$$

where the last inequality holds after substituting k by $n-k$ and applying the result from the first part of the exercise.

C.1-13 Using Stirling's approximation (3.25), we obtain

$$\begin{aligned}\binom{2n}{n} &= \frac{(2n)!}{(n!)^2} \\ &= \frac{\sqrt{4\pi n}\,(2n/e)^{2n}(1+\Theta(1/n))}{2\pi n\,(n/e)^{2n}(1+\Theta(1/n))^2} \\ &= \frac{2^{2n}}{\sqrt{\pi n}} \cdot \frac{1+\Theta(1/n)}{(1+\Theta(1/n))^2}.\end{aligned}$$

Now let's examine the last fraction, which we intentionally didn't reduce because the denominator may not be equal to the square of the numerator. Let c, d be constants such that $c \ge d > 0$, the expression $1+c/n$ bounds the numerator from above, and

the expression $1 + d/n$ bounds the denominator from below. Then

$$\begin{aligned}\frac{1+\Theta(1/n)}{(1+\Theta(1/n))^2} &\le \frac{1+c/n}{(1+d/n)^2}\\ &< \frac{1+c/n}{1+d/n}\\ &= \frac{n+c}{n+d}\\ &= 1+\frac{c-d}{n+d}\\ &< 1+\frac{c-d}{n}\\ &= 1+O(1/n).\end{aligned}$$

Therefore,

$$\binom{2n}{n} = \frac{2^{2n}}{\sqrt{\pi n}}\,(1+O(1/n)).$$

C.1-14 Let's differentiate the binary entropy function H:

$$\begin{aligned}H'(\lambda) &= -\left(\lg\lambda + \lambda\cdot\frac{1}{\lambda\ln 2}\right) - \left(-\lg(1-\lambda) + (1-\lambda)\cdot\frac{-1}{(1-\lambda)\ln 2}\right)\\ &= -\lg\lambda - \lg e + \lg(1-\lambda) + \lg e\\ &= \lg\frac{1-\lambda}{\lambda},\\ H''(\lambda) &= \frac{\lambda}{(1-\lambda)\ln 2}\cdot\frac{-\lambda-(1-\lambda)}{\lambda^2}\\ &= -\frac{\lg e}{\lambda(1-\lambda)}.\end{aligned}$$

The first derivative reaches zero when $\lambda = 1/2$. For $\lambda = 1/2$ the second derivative is negative, so the function H achieves its maximum of $H(1/2) = 1$.

C.1-15 For $n = 0$ the equality holds trivially, so let $n \geq 1$. Then

$$\begin{aligned}\sum_{k=0}^{n} \binom{n}{k} k &= 0 + \sum_{k=1}^{n} \binom{n-1}{k-1} n && \text{(by identity (C.9))} \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} \\ &= n 2^{n-1} && \text{(by indentity (C.4) where } x = y = 1).\end{aligned}$$

C.1-16 We can enumerate the cases with $n = 0, 1, 2, 3$ and $k = 0, 1$, and confirm that the identity holds for all of them.

Now let $n \geq 4$. We have

$$\begin{aligned}\binom{n}{k} &= \frac{n(n-1)\cdots(n-(k-1))}{k!} \\ &= \frac{n^k}{k!}\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right) \\ &\geq \frac{n^k}{k!}\left(1 - \frac{k-1}{n}\right)^{k-1}.\end{aligned}$$

Now consider the function

$$f(x, y) = \left(1 - \frac{x}{n}\right)^y,$$

where $0 \leq x \leq n$ and $y \geq 0$. Observe that the function f monotonically decreases, when x increases for a fixed y, or when y increases for a fixed x. Therefore, since $k - 1 < \sqrt{n}$,

$$\begin{aligned}\left(1 - \frac{k-1}{n}\right)^{k-1} &\geq \left(1 - \frac{\sqrt{n}}{n}\right)^{k-1} \\ &\geq \left(1 - \frac{1}{\sqrt{n}}\right)^{\sqrt{n}}.\end{aligned}$$

It is true that

$$\left(1 - \frac{1}{t}\right)^t \geq \frac{1}{4}$$

for all $t \geq 2$. Therefore, when $n \geq 4$,

$$\begin{aligned}\binom{n}{k} &\geq \frac{n^k}{k!}\left(1 - \frac{1}{\sqrt{n}}\right)^{\sqrt{n}} \\ &\geq \frac{n^k}{4k!}.\end{aligned}$$

C.2 Probability

C.2-1 Let's represent the outcome of the experiment as a 3-string over $\{\text{H}, \text{T}\}$, with the first two elements denoting the results of Professor Rosencrantz's both tosses, and the last element denoting the result of Professor Guildenstern's toss. The sample space consists of $2^3 = 8$ elementary events. The probability that Professor Rosencrantz obtains strictly more heads than Professor Guildenstern is

$$\begin{aligned}\Pr\{\text{HHH}, \text{HHT}, \text{HTT}, \text{THT}\} &= \Pr\{\text{HHH}\} + \Pr\{\text{HHT}\} + \Pr\{\text{HTT}\} + \Pr\{\text{THT}\} \\ &= 1/2.\end{aligned}$$

C.2-2 Let $C_1, C_2, \ldots$ be a finite or countably infinite sequence of events, where

$$C_i = A_i - \bigcup_{j=1}^{i-1} A_j$$

for every $i = 1, 2, \ldots$. Because $\bigcup_i A_i = \bigcup_i C_i$ and any two events C_j, C_k, where $j \neq k$, are mutually exclusive, we get

$$\begin{aligned}\Pr\left\{\bigcup_i A_i\right\} &= \Pr\left\{\bigcup_i C_i\right\} \\ &= \sum_i \Pr\{C_i\} \\ &\leq \sum_i \Pr\{A_i\}.\end{aligned}$$

The last inequality holds, since $\Pr\{C_i\} \leq \Pr\{A_i\}$ for every $i = 1, 2, \ldots$.

C.2-3 Let n_1, n_2, n_3 be the numbers on the selected cards, in order of removing the cards from the deck. Consider the events $A = \{n_1 < n_2\}$ and $B = \{n_2 < n_3\}$. If A occurs, then n_1 could have been chosen in $n_2 - 1$ ways. Similarly, if B occurs, then n_3 can be

chosen in $10 - n_2$ ways. We need to calculate $\Pr\{A \cap B\}$. The number of outcomes in $A \cap B$ is

$$\sum_{n_2=1}^{10} (n_2 - 1)(10 - n_2) = 120.$$

The sample space consists of all possible ordered triples of cards (3-permutations) selected from the deck. By (C.1), the sample space has the size of $10!/7! = 720$. Thus,

$$\begin{aligned}\Pr\{A \cap B\} &= 120/720 \\ &= 1/6.\end{aligned}$$

C.2-4 Since $(A \cap B) \cup (\overline{A} \cap B) = B$, and since events $A \cap B$ and $\overline{A} \cap B$ are mutually exclusive, we get

$$\begin{aligned}\Pr\{A \mid B\} + \Pr\left\{\overline{A} \mid B\right\} &= \frac{\Pr\{A \cap B\}}{\Pr\{B\}} + \frac{\Pr\left\{\overline{A} \cap B\right\}}{\Pr\{B\}} \\ &= \frac{\Pr\{B\}}{\Pr\{B\}} \\ &= 1.\end{aligned}$$

C.2-5 We prove by induction on the number of events n. For $n = 1$ the equality trivially holds, so let's assume $n \geq 2$. We get

$$\begin{aligned}\Pr\left\{\bigcap_{i=1}^{n} A_i\right\} &= \Pr\left\{A_n \cap \bigcap_{i=1}^{n-1} A_i\right\} \\ &= \Pr\left\{\bigcap_{i=1}^{n-1} A_i\right\} \Pr\left\{A_n \;\middle|\; \bigcap_{i=1}^{n-1} A_i\right\} && \text{(by (C.16))} \\ &= \Pr\{A_1\} \Pr\{A_2 \mid A_1\} \Pr\{A_3 \mid A_1 \cap A_2\} \cdots \Pr\left\{A_n \;\middle|\; \bigcap_{i=1}^{n-1} A_i\right\}. \\ & && \text{(by the inductive hypothesis)}\end{aligned}$$

C.2-6 Let $n \geq 3$ and let $S = \{s_{pq} : 1 \leq p, q \leq n\}$ be a sample space with the uniform probability distribution (i.e., $\Pr\left\{s_{pq}\right\} = 1/n^2$ for every $p = 1, 2, \ldots, n$ and $q = 1, 2, \ldots, n$). For each $i = 1, 2, \ldots, n$ consider the event

$$A_i = \{s_{pi} \in S : 1 \leq p \leq i\} \cup \{s_{iq} \in S : i < q \leq n\}.$$

Then $|A_i| = n$, so $\Pr\{A_i\} = 1/n$. Also,

$$|A_{i_1} \cap A_{i_2}| = 1$$

and

$$|A_{i_1} \cap A_{i_2} \cap A_{i_3}| = 0$$

for every distinct indices $1 \le i_1, i_2, i_3 \le n$. Therefore,

$$\begin{aligned}\Pr\{A_{i_1} \cap A_{i_2}\} &= 1/n^2 \\ &= \Pr\{A_{i_1}\} \Pr\{A_{i_2}\},\end{aligned}$$

and so the events $A_1, A_2, \ldots, A_n$ are pairwise independent. On the other hand, for each $k > 2$ and distinct indices $1 \le i_1, \ldots, i_k \le n$,

$$\Pr\left\{\bigcap_{j=1}^{k} A_{i_j}\right\} = 0,$$

while

$$\prod_{j=1}^{k} \Pr\{A_{i_j}\} = 1/n^k,$$

so no k-subset of the events $A_1, A_2, \ldots, A_n$ is mutually independent.

C.2-7 Suppose we have two coins — a fair one and a biased one that always comes up heads. We will randomly pick one of the coins and flip it twice. Let A be the event, that the first flip resulted in heads, let B be the event that the second flip resulted in heads, and let C be the event that the fair coin was picked. Then,

$$\begin{aligned}\Pr\{A\} &= \Pr\{B\} \\ &= (1/2) \cdot (1/2) + (1/2) \cdot 1 \\ &= 3/4, \\ \Pr\{C\} &= 1/2.\end{aligned}$$

The events A and B are not independent, since

$$\begin{aligned}\Pr\{A \cap B\} &= (1/2) \cdot (1/4) + (1/2) \cdot 1 \\ &= 5/8,\end{aligned}$$

while $\Pr\{A\} \Pr\{B\} = 9/16$. But A and B are conditionally independent, given C:

$$\begin{aligned}\Pr\{A \cap B \mid C\} &= 1/4 \\ &= (1/2) \cdot (1/2) \\ &= \Pr\{A \mid C\} \Pr\{B \mid C\}.\end{aligned}$$

C.2-8 Let J, T, and C be the events that Jeff, Tim and Carmine will pass the course, respectively. Furthermore, let G be the event that Professor Gore pointed out that Jeff is the one who will fail. Since the professor can't reveal Carmine's outcome, there must be $\Pr\{G \mid J\} = 0$, $\Pr\{G \mid T\} = 1$, and $\Pr\{G \mid C\} = 1/2$. Before talking to the professor, Carmine knows that his probability of passing is $\Pr\{C\} = 1/3$. His situation after the conversation is described by the event $C \mid G$.

Since $G = (G \cap J) \cup (G \cap T) \cup (G \cap C)$ and since $G \cap J$, $G \cap T$ and $G \cap C$ are mutually exclusive events,

$$\begin{aligned}\Pr\{G\} &= \Pr\{G \cap J\} + \Pr\{G \cap T\} + \Pr\{G \cap C\} \\ &= \Pr\{J\}\Pr\{G \mid J\} + \Pr\{T\}\Pr\{G \mid T\} + \Pr\{C\}\Pr\{G \mid C\}.\end{aligned}$$

Using Bayes's theorem, we obtain

$$\begin{aligned}\Pr\{C \mid G\} &= \frac{\Pr\{C\}\Pr\{G \mid C\}}{\Pr\{G\}} \\ &= \frac{\Pr\{C\}\Pr\{G \mid C\}}{\Pr\{J\}\Pr\{G \mid J\} + \Pr\{T\}\Pr\{G \mid T\} + \Pr\{C\}\Pr\{G \mid C\}} \\ &= \frac{(1/3)\cdot(1/2)}{(1/3)\cdot 0 + (1/3)\cdot 1 + (1/3)\cdot(1/2)} \\ &= 1/3.\end{aligned}$$

Thus, the information that Carmine has obtained does not change his chance of passing.

C.3 Discrete random variables

C.3-1 Define the random variable X to be the sum of the two values showing. Then

$$\begin{aligned}\mathrm{E}[X] &= \sum_{x=2}^{12} x \Pr\{X = x\} \\ &= (1/36)\cdot(2\cdot 1 + 3\cdot 2 + 4\cdot 3 + 5\cdot 4 + 6\cdot 5 + 7\cdot 6 \\ &\qquad + 8\cdot 5 + 9\cdot 4 + 10\cdot 3 + 11\cdot 2 + 12\cdot 1) \\ &= 7.\end{aligned}$$

Define the random variable Y to be the maximum of the values showing. Then

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_{y=1}^{6} y \Pr\{Y = y\} \\
&= (1/36) \cdot (1 \cdot 1 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 7 + 5 \cdot 9 + 6 \cdot 11) \\
&\approx 4.47.
\end{aligned}$$

C.3-2 Let X be the random variable denoting the index of the maximum element in the array $A[1:n]$. For every $i = 1, 2, \ldots, n$, $\Pr\{X = i\} = 1/n$, and therefore

$$\begin{aligned}
\mathrm{E}[X] &= \sum_{i=1}^{n} i \Pr\{X = i\} \\
&= \frac{1}{n} \sum_{i=1}^{n} i \\
&= \frac{1}{n} \cdot \frac{n(n+1)}{2} \\
&= \frac{n+1}{2}.
\end{aligned}$$

In fact, this result is identical for each element of the array, in particular for the maximum element and for the minimum element.

C.3-3 Let X be the random variable representing the player's gain (in dollars) from playing the carnival game once. For $k = 0, 1, 2, 3$, let A_k be the event that the player's number appeared on exactly k dice. Then

$$\begin{aligned}
\Pr\{A_k\} &= \binom{3}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{3-k} \\
&= \binom{3}{k} \frac{5^{3-k}}{6^3}
\end{aligned}$$

We have that

$$\begin{aligned}
\mathrm{E}[X] &= (-1) \cdot \Pr\{A_0\} + 1 \cdot \Pr\{A_1\} + 2 \cdot \Pr\{A_2\} + 3 \cdot \Pr\{A_3\} \\
&= (1/6^3) \cdot ((-1) \cdot 1 \cdot 125 + 1 \cdot 3 \cdot 25 + 2 \cdot 3 \cdot 5 + 3 \cdot 1 \cdot 1) \\
&= -17/216 \\
&\approx -0.08,
\end{aligned}$$

so the player will lose about 8 cents on average in this game.

C.3-4 Consider the subset of the sample space, where $X \geq Y$ holds. Because $Y \geq 0$, then $\mathrm{E}[Y] \geq 0$, and so

$$\begin{aligned}\mathrm{E}[\max\{X, Y\}] &= \mathrm{E}[X] \\ &\leq \mathrm{E}[X] + \mathrm{E}[Y].\end{aligned}$$

The case in which $Y \geq X$ is similar.

C.3-5 According to the definition of independent random variables X and Y, for all x and y, we have

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

Let f and g be any functions over the reals. If $X = x$, then $f(X) = f(x)$, and similarly, if $Y = y$, then $g(Y) = g(y)$. Therefore, we can write the above formula as

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

from which we conclude that $f(X)$ and $g(Y)$ are independent random variables.

C.3-6 For any $t > 0$, we have

$$\begin{aligned}\mathrm{E}[X] &= \sum_{x \geq 0} x \Pr\{X = x\} \\ &= \sum_{0 \leq x < t} x \Pr\{X = x\} + \sum_{x \geq t} x \Pr\{X = x\} \\ &\geq \sum_{x \geq t} x \Pr\{X = x\} \\ &\geq t \sum_{x \geq t} \Pr\{X = x\} \\ &= t \Pr\{X \geq t\}.\end{aligned}$$

C.3-7 Let's pick any real t. Let $A = \{s \in S : X(s) \geq t\}$ and $A' = \{s \in S : X'(s) \geq t\}$. It follows from the relationship between random variables X and X', that if $X'(s) \geq t$,

then $X(s) \geq t$, so $A' \subseteq A$. From the definition of a random variable, we have

$$\begin{aligned}\Pr\{X \geq t\} &= \sum_{s \in A} \Pr\{s\} \\ &= \sum_{s \in A'} \Pr\{s\} + \sum_{s \in A - A'} \Pr\{s\} \\ &\geq \sum_{s \in A'} \Pr\{s\} \\ &= \Pr\left\{X' \geq t\right\}.\end{aligned}$$

C.3-8 Let $f(x) = x^2$ and let $0 \leq \lambda \leq 1$. Then

$$\begin{aligned}f(\lambda x + (1-\lambda)y) - (\lambda f(x) + (1-\lambda) f(y)) \\ &= \lambda^2 x^2 + (1-\lambda)^2 y^2 + 2\lambda(1-\lambda)xy - \lambda x^2 - (1-\lambda)y^2 \\ &= -\lambda(1-\lambda)x^2 - \lambda(1-\lambda)y^2 + 2\lambda(1-\lambda)xy \\ &= -\lambda(1-\lambda)(x-y)^2 \\ &\leq 0,\end{aligned}$$

which means that the function $f(x)$ is convex. By applying it to Jensen's inequality, we obtain

$$\mathrm{E}\left[X^2\right] \geq \mathrm{E}^2[X].$$

C.3-9 Since X takes on only the values 0 and 1, we have $X^2 = X$. Then,

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] \\ &= \mathrm{E}[X] - \mathrm{E}^2[X] \\ &= \mathrm{E}[X](1 - \mathrm{E}[X]) \\ &= \mathrm{E}[X]\,\mathrm{E}[1 - X] \qquad \text{(by linearity of expectation).}\end{aligned}$$

C.3-10 $$\begin{aligned}\mathrm{Var}[aX] &= \mathrm{E}\left[a^2X^2\right] - \mathrm{E}^2[aX] \\ &= a^2\,\mathrm{E}\left[X^2\right] - a^2\,\mathrm{E}^2[X] \qquad \text{(by equation (C.25))} \\ &= a^2\left(\mathrm{E}\left[X^2\right] - \mathrm{E}^2[X]\right) \\ &= a^2\,\mathrm{Var}[X]\end{aligned}$$

C.4 The geometric and binomial distributions

C.4-1 Let X be a random variable representing the number of trials before obtaining a success in a sequence of Bernoulli trials, where a success occurs with probability $p > 0$ and a failure with probability $q = 1 - p$. Then,

$$\begin{aligned}\sum_{k=1}^{\infty} \Pr\{X = k\} &= \sum_{k=1}^{\infty} q^{k-1} p \\ &= p \sum_{k=0}^{\infty} q^k \\ &= \frac{p}{1-q} && \text{(by equation (A.7))} \\ &= 1,\end{aligned}$$

so axiom 2 of the probability axioms holds.

C.4-2 We have a success when of all six coins, any three came up heads and the other three came up tails — it is therefore $\binom{6}{3} = 20$ ways to obtain a success. There are $2^6 = 64$ possible outcomes, so the probability of success is $p = 20/64 = 5/16$. By (C.36), before obtaining one we need to make $1/p \approx 3.2$ flips on average.

C.4-3 Let X be a random variable having a geometric distribution with the probability $p > 0$ of a success and the probability $q = 1 - p$ of a failure. First, let's find the expectation of X^2:

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= \sum_{k=1}^{\infty} k^2 \Pr\{X = k\} \\ &= \sum_{k=1}^{\infty} k^2 q^{k-1} p \\ &= \frac{p}{q} \sum_{k=0}^{\infty} k^2 q^k \\ &= \frac{p}{q} \cdot \frac{q(1+q)}{(1-q)^3} && \text{(by Exercise A.1-6)} \\ &= (1+q)/p^2.\end{aligned}$$

From the definition (C.31) of variance and equation (C.36), we have

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}\left[X^2\right] - \mathrm{E}^2[X] \\ &= (1+q)/p^2 - 1/p^2 \\ &= q/p^2.\end{aligned}$$

C.4-4

$$\begin{aligned}b(k;n,p) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \binom{n}{n-k} q^{n-k} (1-q)^k && \text{(by equation (C.3))} \\ &= b(n-k;n,q)\end{aligned}$$

C.4-5

The binomial distribution achieves a maximum at an integer k, such that $np - q \le k \le (n+1)p$, so a good approximation of the maximum is the value for $k = np$. Since np may not be integer, we'll sacrifice mathematical rigor in order to simplify our calculations:

$$\begin{aligned}b(np;n,p) &= \binom{n}{np} p^{np} (1-p)^{n-np} \\ &= \frac{n!}{(np)!\,(n-np)!}\, p^{np} (1-p)^{n-np} \\ &= \frac{n!}{(np)!\,(nq)!}\, p^{np} q^{nq}.\end{aligned}$$

Using Stirling's approximation (3.25), we can simplify the first factor of the last expression above:

$$\begin{aligned}\frac{n!}{(np)!\,(nq)!} &\approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi np}\left(\frac{np}{e}\right)^{np}\sqrt{2\pi nq}\left(\frac{nq}{e}\right)^{nq}} \\ &= \frac{\left(\frac{n}{e}\right)^n \left(\frac{e}{np}\right)^{np} \left(\frac{e}{nq}\right)^{nq}}{\sqrt{2\pi npq}} \\ &= \frac{1}{p^{np} q^{nq} \sqrt{2\pi npq}}.\end{aligned}$$

Hence, the value of the maximum of $b(k;n,p)$ is roughly equal to $1/\sqrt{2\pi npq}$.

C.4-6 We are interested in the value of $b(0; n, 1/n)$ and we can approximate it by examining its limit:

$$\begin{aligned}\lim_{n\to\infty} b(0; n, 1/n) &= \lim_{n\to\infty} (1 - 1/n)^n \\ &= 1/e \qquad \text{(by equation (3.16)).}\end{aligned}$$

In the second part we similarly approximate $b(1; n, 1/n)$:

$$\begin{aligned}\lim_{n\to\infty} b(1; n, 1/n) &= \lim_{n\to\infty} \frac{(1 - 1/n)^n}{1 - 1/n} \\ &= \frac{1/e}{1} \qquad \text{(by equation (3.16))} \\ &= 1/e.\end{aligned}$$

C.4-7 We'll calculate the probability that the professors get the same number of heads in two ways. In the first one, we'll treat the result of each experiment as a $2n$-element sequence of heads and tails, such that its initial n terms are the results obtained by Professor Rosencrantz, and the final n terms are the results obtained by Professor Guildenstern. There are $2^{2n} = 4^n$ of all such sequences. Following the hint, we call a head a success for Professor Rosencrantz and we call a tail a success for Professor Guildenstern. Note that both professors obtain the same number of heads if and only if they achieve exactly n successes in total. The number of ways this can be done is the number of choices of n items responsible for successes among all $2n$ items in the sequence. This number is equal to $\binom{2n}{n}$, so the probability we are looking for is $\binom{2n}{n}/4^n$.

Another approach to determine the desired probability, is to define random variables R and G denoting the number of heads obtained by Professor Rosencrantz and by Professor Guildenstern, respectively. Again, defining successes for each professor according to the hint, we get that R has the binomial distribution $b(k; n, 1/2)$ and G has the binomial distribution $b(n - k; n, 1/2)$. The events $R = k$ and $G = k$ are

independent, so the probability we are looking for, is

$$\begin{aligned}\sum_{k=0}^{n} \Pr\{R = k \text{ and } G = k\} &= \sum_{k=0}^{n} \Pr\{R = k\} \Pr\{G = k\} \\ &= \sum_{k=0}^{n} b(k; n, 1/2)\, b(n-k; n, 1/2) \\ &= \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{n-k} \binom{n}{n-k} \left(\frac{1}{2}\right)^{n-k} \left(\frac{1}{2}\right)^{k} \\ &= \frac{\sum_{k=0}^{n} \binom{n}{k}^2}{4^n}.\end{aligned}$$

Comparing the results obtained in both ways, we get the identity

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

C.4-8 Let $0 \le \lambda \le 1$. Using the inequality

$$\binom{n}{\lambda n} \le 2^{nH(\lambda)},$$

that follows from (C.7), we obtain

$$\begin{aligned}b(k; n, 1/2) &= \binom{n}{k} \left(\frac{1}{2}\right)^{k} \left(\frac{1}{2}\right)^{n-k} \\ &= \binom{n}{(k/n)\, n} \left(\frac{1}{2}\right)^{n} \\ &\le \frac{2^{nH(k/n)}}{2^n} \\ &= 2^{nH(k/n)-n}.\end{aligned}$$

C.4-9 Let X' be the random variable denoting the number of successes in a series of Bernoulli trials, each with a probability p of success. Using the result from Exercise C.4-10, where $p'_i = p$ for every $i = 1, 2, \ldots, n$, gives

$$\Pr\{X \ge k\} \le \Pr\left\{X' \ge k\right\}.$$

The random variable X' has a binomial distribution, so the above inequality can be reformulated as

$$\Pr\{X \geq k\} \leq \sum_{i=k}^{n} b(i; n, p).$$

Then,

$$\begin{aligned}\Pr\{X < k\} &= 1 - \Pr\{X \geq k\} \\ &\geq \sum_{i=0}^{n} b(i; n, p) - \sum_{i=k}^{n} b(i; n, p) \qquad \text{(by equation (C.40))} \\ &= \sum_{i=0}^{k-1} b(i; n, p).\end{aligned}$$

C.4-10 Let S be the sample space containing all possible sequences of outcomes during n Bernoulli trials. A set A corresponds to a sequence $s \in S$, and $X(s)$ is the number of successes in s.

We'll show how to obtain a set A' of Bernoulli trials by performing experiments on the trials of A. Let A_i denote the event that the ith trial of A is a success, and let A'_i be the event that the ith trial of A' is a success. If A_i occurs, we'll make A'_i to also occur. Otherwise, a success in the ith trial of A' will be generated with some probability r_i. Then,

$$\begin{aligned}p'_i &= \Pr\{A'_i\} \\ &= \Pr\{A'_i \cap A_i\} + \Pr\{A'_i \cap \overline{A_i}\} \\ &= p_i + (1 - p_i) \cdot r_i,\end{aligned}$$

and so,

$$r_i = \frac{p'_i - p_i}{1 - p_i}.$$

Reusing the sample space S, we'll denote by $X'(s)$ the number of successes in a series of the n trials obtained by following the above procedure, based on the outcomes in s. For any $s \in S$ and its corresponding set A, it's clear that after constructing the set A' this way, it's impossible that there are fewer successes in A' than in the original set A. In other words, $X'(s) \geq X(s)$ holds. Using Exercise C.3-7, we get the desired result.

C.5 The tails of the binomial distribution

C.5-1 Intuitively, the situation when each flip in a series results in a head is less likely when the series consists of twice more flips, while we expect the same number of heads as before. Let's prove our intuition. Let X_m for $m \geq 0$ be the random variable denoting the number of heads obtained in m flips of a fair coin. When $n \leq m$, $\Pr\{X_m = n\} = b(n; m, 1/2)$. Thus, the ratio of both studied probabilities is

$$\begin{aligned}
\frac{\Pr\{X_{2n} = n\}}{\Pr\{X_n = n\}} &= \frac{\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}}{\binom{n}{n}\left(\frac{1}{2}\right)^{n}} \\
&= \frac{\binom{2n}{n}}{2^n} \\
&= \frac{2^n}{\sqrt{\pi n}}\left(1 + O(1/n)\right) && \text{(by equation (C.11))} \\
&> 1,
\end{aligned}$$

which confirms our guess.

C.5-2

Proof of Corollary C.6 Let Y be the random variable denoting the total number of failures in this experiment. Then, $\Pr\{X > k\} = \Pr\{Y < n - k\}$. Since $np < k < n$, or $0 < n - k < nq$, we can apply Theorem C.4 for Y, after inverting the roles of successes and failures, to obtain the desired inequality. ■

Proof of Corollary C.7 Similarly as in the previous proof, let's treat the probability of getting more than k successes as the probability of getting fewer than $n - k$ failures. Since $(np + n)/2 < k < n$, we have $0 < n - k < nq/2$, and after swapping the roles of successes and failures, the result follows from Corollary C.5. ■

C.5-3 Let $p = a/(a+1)$ and $q = 1 - p = 1/(a+1)$, so that $a = p/q$. Then,

$$\begin{aligned}\sum_{i=0}^{k-1} \binom{n}{i} a^i &= \sum_{i=0}^{k-1} \binom{n}{i} \left(\frac{p}{q}\right)^i \\ &= \frac{\sum_{i=0}^{k-1} \binom{n}{i} p^i q^{n-i}}{q^n} \\ &= \frac{\sum_{i=0}^{k-1} b(i; n, p)}{q^n},\end{aligned}$$

so by applying Theorem C.4 to the last summation, we obtain

$$\begin{aligned}\sum_{i=0}^{k-1} \binom{n}{i} a^i &< \frac{kq}{q^n(np-k)}\, b(k; n, p) \\ &= \frac{k/(a+1)}{(1/(a+1))^n(na/(a+1)-k)}\, b(k; n, a/(a+1)) \\ &= (a+1)^n \frac{k}{na - k(a+1)}\, b(k; n, a/(a+1)).\end{aligned}$$

C.5-4 Using the fact that $\binom{n}{i} \geq 1$ for all $i = 0, 1, \ldots, n$, we have

$$\begin{aligned}\sum_{i=0}^{k-1} p^i q^{n-i} &\leq \sum_{i=0}^{k-1} \binom{n}{i} p^i q^{n-i} \\ &= \sum_{i=0}^{k-1} b(i; n, p) \\ &< \frac{kq}{np-k}\, b(k; n, p) && \text{(by Theorem C.4)} \\ &\leq \frac{kq}{np-k} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} && \text{(by Lemma C.1).}\end{aligned}$$

C.5-5 Let $Y = n - X$ and let $\nu = \mathrm{E}[Y]$. By linearity of expectation, $\nu = n - \mathrm{E}[X] = n - \mu$. Since $r > \nu$, we can use Theorem C.8 to obtain

$$\begin{aligned}\Pr\{\mu - X \geq r\} &= \Pr\{n - \nu - X \geq r\} \\ &= \Pr\{Y - \nu \geq r\} \\ &\leq \left(\frac{\nu e}{r}\right)^r \\ &= \left(\frac{(n-\mu)e}{r}\right)^r .\end{aligned}$$

Similarly, in the second part, if we let $Y = n - X$, then $\mathrm{E}[Y] = n - \mathrm{E}[X] = n - np = nq$. Since $r > nq$, we use Corollary C.9 to get

$$\begin{aligned}\Pr\{np - X \geq r\} &= \Pr\{n - nq - X \geq r\} \\ &= \Pr\{Y - nq \geq r\} \\ &\leq \left(\frac{nqe}{r}\right)^r .\end{aligned}$$

C.5-6

Lemma
Let α, p, and q be nonnegative reals, such that $p + q = 1$. Then,

$$pe^{\alpha q} + qe^{-\alpha p} \leq e^{\alpha^2/2}. \tag{3}$$

Proof Consider the function $f(\alpha) = e^{\alpha^2/2} - (pe^{\alpha q} + qe^{-\alpha p})$ for $\alpha \geq 0$, and examine its derivatives:

$$\begin{aligned}f'(\alpha) &= \alpha e^{\alpha^2/2} - pq(e^{\alpha q} - e^{-\alpha p}), \\ f''(\alpha) &= \alpha^2 e^{\alpha^2/2} + e^{\alpha^2/2} - pq(qe^{\alpha q} + pe^{-\alpha p}).\end{aligned}$$

It's true that $f''(0) = 1 - pq \geq 0$ and we'll show that $f''(\alpha) > 0$ for all $\alpha > 0$. Since

$$\begin{aligned}pq - 1/4 &= p(1-p) - 1/4 \\ &= -p^2 + p - 1/4 \\ &= -(p - 1/2)^2 \\ &\leq 0\end{aligned}$$

or, equivalently, since $pq \le 1/4$,

$$\begin{aligned} pq(qe^{\alpha q} + pe^{-\alpha p}) &\le (qe^{\alpha q} + pe^{-\alpha p})/4 \\ &\le (e^{\alpha q} + e^{-\alpha p})/4 \\ &= e^{-\alpha p}(e^{\alpha q+\alpha p} + 1)/4 \\ &= e^{-\alpha p}(e^{\alpha} + 1)/4 \\ &\le (e^{\alpha} + 1)/4. \end{aligned}$$

Combining this inequality with $\alpha^2 e^{\alpha^2/2} > 0$, that holds for all $\alpha > 0$, we get

$$f''(\alpha) > e^{\alpha^2/2} - (e^{\alpha} + 1)/4.$$

Now it suffices to show that $4e^{\alpha^2/2} \ge e^{\alpha} + 1$ or, since $e^{\alpha^2/2} > 1$, that

$$3e^{\alpha^2/2} \ge e^{\alpha}.$$

This is equivalent to $3e^{\alpha^2/2}/e^{\alpha} = 3e^{\alpha^2/2-\alpha} \ge 1$, which holds when

$$\alpha^2/2 - \alpha + \ln 3 \ge 0.$$

The left-hand side expression attains the minimum at $\alpha = 1$, equal to $\ln 3 - 1/2 > 0$, which concludes the proof that $f''(\alpha) \ge 0$ for $\alpha \ge 0$.

The fact we've just shown implies that the function $f'(\alpha)$ is monotonically increasing for all $\alpha \ge 0$. Since $f'(0) = 0$, it must be $f'(\alpha) \ge 0$ for all $\alpha \ge 0$, which in turn implies that the function $f(\alpha)$ is monotonically increasing. And since $f(0) = 0$, it holds that $f(\alpha) \ge 0$ for all $\alpha \ge 0$, which completes the proof. ■

The beginning of the main reasoning is based on the proof of Theorem C.8:

$$\begin{aligned} \Pr\{X - \mu \ge r\} &= \Pr\left\{e^{\alpha(X-\mu)} \ge e^{\alpha r}\right\} \\ &\le \mathrm{E}\left[e^{\alpha(X-\mu)}\right] e^{-\alpha r}. \end{aligned}$$

Then, with the same notations as in the original proof, it follows that

$$\mathrm{E}\left[e^{\alpha(X-\mu)}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{\alpha(X_i-p_i)}\right].$$

Using inequality (3) for $p = p_i$ and $q = q_i$, we get

$$\begin{aligned} \mathrm{E}\left[e^{\alpha(X_i-p_i)}\right] &= e^{\alpha(1-p_i)} p_i + e^{\alpha(0-p_i)} q_i \\ &= p_i e^{\alpha q_i} + q_i e^{-\alpha p_i} \\ &\le e^{\alpha^2/2}, \end{aligned}$$

and so

$$\begin{aligned}\mathrm{E}\left[e^{\alpha(X-\mu)}\right] &= \prod_{i=1}^{n} \mathrm{E}\left[e^{\alpha(X_i-p_i)}\right] \\ &\le \prod_{i=1}^{n} e^{\alpha^2/2} \\ &= \exp(\alpha^2 n/2).\end{aligned}$$

Therefore,

$$\begin{aligned}\Pr\{X-\mu \ge r\} &\le \mathrm{E}\left[e^{\alpha(X-\mu)}\right] e^{-\alpha r} \\ &\le \exp(\alpha^2 n/2 - \alpha r).\end{aligned}$$

Now we need to choose the value of α that minimizes the last expression. The argument of the exponential function in that expression is a quadratic function with respect to α, so it's easy to show that it attains a minimum at $\alpha = r/n$. We finally get

$$\begin{aligned}\Pr\{X-\mu \ge r\} &\le \exp\left((r/n)^2 n/2 - (r/n)r\right) \\ &= e^{-r^2/2n}.\end{aligned}$$

C.5-7 The right-hand side of inequality (C.51) can be expressed as the function $f(\alpha) = \exp(\mu e^{\alpha} - \alpha r)$, where $\alpha > 0$ and $r > \mu > 0$. We examine its derivatives:

$$\begin{aligned}f'(\alpha) &= (\mu e^{\alpha} - r)\exp(\mu e^{\alpha} - \alpha r), \\ f''(\alpha) &= \left(\mu e^{\alpha} + (\mu e^{\alpha} - r)^2\right)\exp(\mu e^{\alpha} - \alpha r).\end{aligned}$$

$f'(\alpha) = 0$ for $\alpha = \ln(r/\mu)$, and $f''(\alpha) > 0$ for all $\alpha > 0$. Therefore, $f(\alpha)$ is minimized at $\alpha = \ln(r/\mu)$.

Problems

C-1 ***The Monty Hall problem***

a. Let A be the event that the door we chose at first is the one with the automobile, and let W be the event of winning the automobile. Then $\Pr\{A\} = 1/3$ and

$$\begin{aligned}\Pr\{W\} &= \Pr\{W \cap A\} + \Pr\left\{W \cap \overline{A}\right\} \\ &= \Pr\{A\}\Pr\{W \mid A\} + \Pr\left\{\overline{A}\right\}\Pr\left\{W \mid \overline{A}\right\} \qquad \text{(by equation (C.16))} \\ &= (1/3)\cdot\Pr\{W \mid A\} + (2/3)\cdot\Pr\left\{W \mid \overline{A}\right\}.\end{aligned}$$

Let's calculate the value of the above probability depending on the decision we made after Carol has revealed the first goat. If we decided to stick to our current choice, $\Pr\{W \mid A\} = 1$ and $\Pr\{W \mid \overline{A}\} = 0$, so we win with probability $\Pr\{W\} = 1/3$. Otherwise, we switched, and then $\Pr\{W \mid A\} = 0$ and $\Pr\{W \mid \overline{A}\} = 1$. In this case the probability of winning is $\Pr\{W\} = 2/3$. Thus, by choosing to switch, we double the chances of winning the prize.

b. In this game we make two decisions — first we choose one of the three doors, and then, after one of the doors have been opened, whether to stick or switch. There are therefore six outcomes, of which three correspond to winning the game. The results are detailed in Figure C-1.

	stick	switch
the right door	$(1 - p_{\text{right}}\,p_{\text{switch}})/3$	$p_{\text{right}}\,p_{\text{switch}}/3$
the first wrong door	$(1 - p_{\text{wrong}}\,p_{\text{switch}})/3$	$p_{\text{wrong}}\,p_{\text{switch}}/3$
the second wrong door	$(1 - p_{\text{wrong}}\,p_{\text{switch}})/3$	$p_{\text{wrong}}\,p_{\text{switch}}/3$

Figure C-1 The probability distribution in the Monty Hall's game. Rows represent our first choice and columns represent our second choice, and each cell contains the probability of the corresponding outcome. The cells shaded blue correspond to the winning outcomes.

c. By summing up the probabilities of the winning outcomes from the table in Figure C-1, we get

$$\begin{aligned} p_{\text{win}} &= (1 - p_{\text{right}}\,p_{\text{switch}})/3 + p_{\text{wrong}}\,p_{\text{switch}}/3 + p_{\text{wrong}}\,p_{\text{switch}}/3 \\ &= \frac{1}{3}(2\,p_{\text{wrong}}\,p_{\text{switch}} - p_{\text{right}}\,p_{\text{switch}} + 1). \end{aligned}$$

d. In order to minimize our chances p_{win} of winning, given $p_{\text{switch}} > 0$, Monty needs to minimize the term $2\,p_{\text{wrong}}\,p_{\text{switch}}$, as well as maximize the term $p_{\text{right}}\,p_{\text{switch}}$. Therefore, his best strategy is $p_{\text{right}} = 1$ and $p_{\text{wrong}} = 0$. With such values, our chances of winning are

$$\begin{aligned} p_{\text{win}} &= (1 - p_{\text{switch}})/3 \\ &< 1/3. \end{aligned}$$

e. If $p_{\text{switch}} = 0$, the winning probability reduces to $p_{\text{win}} = 1/3$, independently of any choice for p_{right} and p_{wrong}.

f. In order to maximize p_{win} for fixed p_{right} and p_{wrong}, we need to maximize the expression

$$2p_{\text{wrong}}p_{\text{switch}} - p_{\text{right}}p_{\text{switch}} = p_{\text{switch}}(2p_{\text{wrong}} - p_{\text{right}}). \tag{4}$$

If $p_{\text{right}} \le 2p_{\text{wrong}}$, we just set p_{switch} to the maximum possible value of 1. Otherwise, the value of (4) is nonpositive, so by letting $p_{\text{switch}} = 0$ we make it achieve 0.

g. As a function of p_{right} and p_{wrong}, the expression (4) is minimized, when $p_{\text{right}} = 1$ and $p_{\text{wrong}} = 0$. Then, $p_{\text{win}} = (1 - p_{\text{switch}})/3$, so our best strategy to maximize p_{win} is to choose $p_{\text{switch}} = 0$.

h. Let M be the event that Monty gives us an opportunity to switch, and let A and W be the events of the same meaning as in the solution to part (a). Then, $\Pr\{M \mid A\} = p_{\text{right}}$ and $\Pr\{M \mid \overline{A}\} = p_{\text{wrong}}$, and so

$$\begin{aligned}\Pr\{M\} &= \Pr\{A\}\Pr\{M \mid A\} + \Pr\{\overline{A}\}\Pr\{M \mid \overline{A}\} \\ &= (1/3)p_{\text{right}} + (2/3)p_{\text{wrong}}.\end{aligned}$$

Now let S be the event that we switched. Then,

$$\begin{aligned}W \cap M &= \big((A \cap \overline{S}) \cup (\overline{A} \cap S)\big) \cap M \\ &= \big((A \cap M) \cap \overline{S}\big) \cup \big((\overline{A} \cap M) \cap S\big),\end{aligned}$$

and the events $A \cap M$ and $\overline{S}$ are independent, and so are the events $\overline{A} \cap M$ and S. Thus, by the definition (C.16) of conditional probability,

$$\begin{aligned}\Pr\{W \mid M\} &= \frac{\Pr\{W \cap M\}}{\Pr\{M\}} \\ &= \frac{\Pr\{(A \cap M) \cap \overline{S}\} + \Pr\{(\overline{A} \cap M) \cap S\}}{\Pr\{M\}} \\ &= \frac{\Pr\{A\}\Pr\{M \mid A\}\Pr\{\overline{S}\} + \Pr\{\overline{A}\}\Pr\{M \mid \overline{A}\}\Pr\{S\}}{\Pr\{M\}} \\ &= \frac{(1/3)p_{\text{right}}(1 - p_{\text{switch}}) + (2/3)p_{\text{wrong}}p_{\text{switch}}}{(1/3)p_{\text{right}} + (2/3)p_{\text{wrong}}} \\ &= \frac{p_{\text{right}} - p_{\text{right}}p_{\text{switch}} + 2p_{\text{wrong}}p_{\text{switch}}}{p_{\text{right}} + 2p_{\text{wrong}}}.\end{aligned}$$

The conditional probability $\Pr\{W \mid M\}$ is defined whenever $\Pr\{M\} \ne 0$, which holds when $p_{\text{right}} + 2p_{\text{wrong}} \ne 0$ or, equivalently, when $p_{\text{right}} \ne 0$ or $p_{\text{wrong}} \ne 0$.

i. The value of the expression for $p_{\text{switch}} = 1/2$ is

$$\begin{aligned}\Pr\{W \mid M\} &= \frac{p_{\text{right}}/2 + p_{\text{wrong}}}{p_{\text{right}} + 2p_{\text{wrong}}} \\ &= 1/2.\end{aligned}$$

When $p_{\text{switch}} < 1/2$, Monty can select $p_{\text{right}} = 0$ and any $p_{\text{wrong}} > 0$, so that

$$\begin{aligned}\Pr\{W \mid M\} &= \frac{2p_{\text{wrong}}p_{\text{switch}}}{2p_{\text{wrong}}} \\ &= p_{\text{switch}} \\ &< 1/2.\end{aligned}$$

Similarly, when $p_{\text{switch}} > 1/2$, Monty's best strategy is to choose any $p_{\text{right}} > 0$ and $p_{\text{wrong}} = 0$, in which case

$$\begin{aligned}\Pr\{W \mid M\} &= \frac{p_{\text{right}} - p_{\text{right}}p_{\text{switch}}}{p_{\text{right}}} \\ &= 1 - p_{\text{switch}} \\ &< 1/2.\end{aligned}$$

j. In part (i) we showed that for any value of p_{switch} other than $1/2$, Monty can always find a strategy that will reduce our chances of winning, as compared to the situation when we choose $p_{\text{switch}} = 1/2$.

In this problem we studied different strategies for Monty and for the player. The analysis of the original problem showed that switching can actually double the player's chances of winning the game, which for most people may seem counterintuitive. We have seen that for any Monty's strategy, there is an optimal strategy that the player can follow in order to maximize their chances to win, given that they know the Monty's plan. And vice versa, knowing how the player will act, Monty can always reduce the player's chances.

C-2 *Balls and bins*

a. We place each ball in one of the b bins. There are b ways to choose a bin for each ball. Thus, we can make n such choices in b^n ways.

b. We can view this problem as counting all possible arrangements of n distinguishable balls and $b - 1$ indistinguishable sticks. Informally speaking, the sticks in such arrangements partition the balls into subsequences of balls that end up in different bins.

There are $(b + n - 1)!$ of all such arrangements, but since we do not distinguish sticks, we divide this number by the number of permutations of the sticks, $(b - 1)!$, to obtain the number $\frac{(b+n-1)!}{(b-1)!}$ of ways to place the balls in the bins in this variant.

c. Compared to the situation from part (b), here we are not distinguishing the balls, so every permutation of them — with the positions of the sticks unchanged — represents the same arrangement of the balls in the bins. Thus, we have $\frac{(b+n-1)!}{n!\,(b-1)!} = \binom{b+n-1}{n}$ ways to place the balls.

d. Out of b bins we choose n that will contain a ball. There are $\binom{b}{n}$ ways to do this.

e. First, we place a ball in each bin, so that none of the bins are empty. By part (c), we can place the remaining $n - b$ balls in the b bins in $\binom{b+(n-b)-1}{n-b} = \binom{n-1}{n-b} = \binom{n-1}{b-1}$ ways.

D Matrices

D.1 Matrices and matrix operations

D.1-1 By the definition of a symmetric matrix, $A = A^{\mathrm{T}}$ and $B = B^{\mathrm{T}}$ or, equivalently, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$, $a_{ij} = a_{ji}$ and $b_{ij} = b_{ji}$. Let $C = A + B$ and $D = A - B$. Then for all $i = 1, 2, \ldots, n$ and all $j = 1, 2, \ldots, n$,

$$\begin{aligned} c_{ij} &= a_{ij} + b_{ij} \\ &= a_{ji} + b_{ji} \\ &= c_{ji}, \end{aligned}$$

and, similarly,

$$\begin{aligned} d_{ij} &= a_{ij} - b_{ij} \\ &= a_{ji} - b_{ji} \\ &= d_{ji}, \end{aligned}$$

so C and D are symmetric.

D.1-2 Suppose that A is a $p \times q$ matrix and B is a $q \times r$ matrix. We'll denote $A^{\mathrm{T}} = (a_{ij}^{\mathrm{T}}) = (a_{ji})$ and $B^{\mathrm{T}} = (b_{ij}^{\mathrm{T}}) = (b_{ji})$. Let $C = (AB)^{\mathrm{T}}$. Then, C is an $r \times p$ matrix, such that for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, r$,

$$\begin{aligned} c_{ji} &= \sum_{k=1}^{q} a_{ik} b_{kj} \\ &= \sum_{k=1}^{q} b_{jk}^{\mathrm{T}} a_{ki}^{\mathrm{T}}, \end{aligned}$$

so $C = B^{\mathrm{T}} A^{\mathrm{T}}$.

This fact implicates that

$$\begin{aligned}(A^{\mathrm{T}}A)^{\mathrm{T}} &= A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} \\ &= A^{\mathrm{T}}A,\end{aligned}$$

and so the matrix $A^{\mathrm{T}}A$ is symmetric.

D.1-3 Let $L' = (l'_{ij})$ and $L'' = (l''_{ij})$ be $n \times n$ lower-triangular matrices. Let $L = L'L''$. For $1 \le i < j \le n$,

$$\begin{aligned}l_{ij} &= \sum_{k=1}^{n} l'_{ik} l''_{kj} \\ &= \sum_{k=1}^{j-1} l'_{ik} l''_{kj} + \sum_{k=j}^{n} l'_{ik} l''_{kj} \\ &= \sum_{k=1}^{j-1} l'_{ik} \cdot 0 + \sum_{k=j}^{n} 0 \cdot l''_{kj} \\ &= 0,\end{aligned}$$

so L is lower-triangular.

D.1-4 Let $PA = A' = (a'_{ij})$. Suppose that for $i = 1, 2, \ldots, n$, row i of the matrix P has 1 in column c_i (i.e., $p_{ic_i} = 1$). Then,

$$\begin{aligned}a'_{ij} &= \sum_{k=1}^{n} p_{ik} a_{kj} \\ &= p_{ic_i} a_{c_i j} \\ &= a_{c_i j}.\end{aligned}$$

Therefore, the element a'_{ij} in the matrix PA is an element from A from the same row but possibly from a different column. This means that PA is formed by permuting rows of A.

Similarly, let $AP = A'' = (a''_{ij})$ and for $j = 1, 2, \ldots, n$ let r_j be the row number, such that $p_{r_j j} = 1$. Then,

$$\begin{aligned}a''_{ij} &= \sum_{k=1}^{n} a_{ik} p_{kj} \\ &= a_{ir_j} p_{r_j j} \\ &= a_{ir_j},\end{aligned}$$

which implies that AP is A with its columns permuted.

Now suppose that A is a permutation matrix. Permuting rows of A preserves the property of permutation matrices that each row and each column has exactly one 1, and 0s elsewhere. Therefore, PA is also a permutation matrix.

D.2 Basic matrix properties

D.2-1

$$\begin{aligned} B &= BI \\ &= B(AC) \\ &= (BA)C \\ &= IC \\ &= C \end{aligned}$$

D.2-2

Let L be an $n \times n$ lower-triangular matrix. We'll prove by induction on n that the determinant of L is the product of its diagonal elements. The base case is when $n = 1$, and from the definition of determinant we have $\det(L) = l_{11}$. Now suppose that $n \geq 2$. The minor $L_{[11]}$ is a lower-triangular matrix, so let's assume as an inductive hypothesis, that $\det(L_{[11]}) = \prod_{i=2}^{n} l_{ii}$. Then we have

$$\begin{aligned} \det(L) &= \sum_{j=1}^{n} (-1)^{1+j} l_{1j} \det(L_{[1j]}) \\ &= (-1)^{1+1} l_{11} \det(L_{[11]}) \\ &= l_{11} \prod_{i=2}^{n} l_{ii} \\ &= \prod_{i=1}^{n} l_{ii}, \end{aligned} \tag{5}$$

which concludes the inductive proof.

Now let U be an $n \times n$ upper-triangular matrix. There exists a lower-triangular

matrix L, such that $U = L^{\mathrm{T}}$. Then U and L have the same diagonal elements, and so

$$\begin{aligned}\det(U) &= \det(L^{\mathrm{T}}) \\ &= \det(L) && \text{(by Theorem D.4)} \\ &= \prod_{i=1}^{n} l_{ii} && \text{(by equation (5))} \\ &= \prod_{i=1}^{n} u_{ii}.\end{aligned}$$

Now let L be a nonsingular $n \times n$ lower-triangular matrix and let $L' = (l'_{ij})$ be the inverse of L. We use induction to show that L' is a lower-triangular matrix. When $n = 1$, L' is a 1×1 matrix, which is trivially lower-triangular. For the inductive step, let $n \geq 2$ and assume that the minor $L'_{[11]}$ itself is lower-triangular. For $j = 1, 2, \ldots, n$, the element in the first row and the jth column of I_n is equal to $\sum_{k=1}^{n} l_{1k} l'_{kj} = l_{11} l'_{1j}$. If $j = 1$, then $l_{11} l'_{11} = 1$, so $l'_{11} = 1/l_{11} \neq 0$. Note that the existence of the inverse of L implies $l_{11} \neq 0$. If $j > 1$, $l_{11} l'_{1j} = 0$, so l'_{1j} has to be 0. Combining these results with the inductive hypothesis, we have that L' is lower-triangular.

D.2-3 Suppose that P is an $n \times n$ matrix. We'll prove that P is invertible by showing that $P^{\mathrm{T}} = (p^{\mathrm{T}}_{ij})$ is its inverse. Suppose that for $i = 1, 2, \ldots, n$, c_i is the number of column, such that $p_{ic_i} = 1$. Then $p^{\mathrm{T}}_{c_i i} = 1$ is the only nonzero element in column i of P^{T}.

If we let $A = PP^{\mathrm{T}}$, then for every $1 \leq i \leq n$ and $1 \leq j \leq n$, we have that

$$\begin{aligned}a_{ij} &= \sum_{k=1}^{n} p_{ik} p^{\mathrm{T}}_{kj} \\ &= \sum_{k=1}^{n} p_{ik} p_{jk} \\ &= p_{ic_i} p_{jc_i} \\ &= p_{jc_i}.\end{aligned}$$

By the fact that P is a permutation matrix, we have that if $i = j$, then $a_{ij} = 1$, otherwise $a_{ij} = 0$. That is, $A = I_n$, and so $P^{\mathrm{T}} = P^{-1}$.

In a permutation matrix P each row and each column has exactly one 1, and 0s elsewhere. The rows of P are the columns of P^{T}, and the columns of P are the rows of P^{T}, so the permutation matrix property is preserved in P^{T}.

D.2-4 Let C be the $n \times n$ matrix with $c_{ij} = 1$, and 0s elsewhere. Observe that CA is the matrix, in which the ith row is the jth row of A, while other rows consist of 0s.

Similarly, BC is the matrix, in which the jth column is the ith column of B, while other columns consist of 0s. Let $D = I + C$ and $D' = I - C$. Then, $A' = DA$ and $B' = BD'$.

Observe, that since $i \neq j$, for any $1 \leq p, q \leq n$ the terms c_{pk} and c_{kq} can't be both nonzero. Thus, $\sum_{k=1}^{n} c_{pk}c_{kq} = 0$, and so CC is a zero matrix. Then,

$$\begin{aligned} DD' &= (I + C)(I - C) \\ &= II + CI - IC - CC \\ &= I + C - C - CC \\ &= I, \end{aligned}$$

which means that $D' = D^{-1}$.

Returning to the matrices A' and B', we get

$$\begin{aligned} A'B' &= DABD' \\ &= D(AB)D^{-1} \\ &= (DI)D^{-1} \\ &= DD^{-1} \\ &= I, \end{aligned}$$

hence $B' = (A')^{-1}$.

D.2-5 We need only prove the forward direction, since the inverse of A^{-1} is $(A^{-1})^{-1} = A$ and the backward direction is symmetric. Suppose that every entry of $A^{-1} = (a_{ij}^{-1})$ is real. We'll use induction to show that every entry of A is real.

If $n = 1$, A^{-1} has a single real entry a_{11}^{-1}, and the only entry of A is $a_{11} = 1/a_{11}^{-1}$, so it is real. Now let $n \geq 2$ and assume that every entry of $A_{[11]}$ is real. Suppose that there exists $2 \leq i \leq n$, such that a_{i1} is complex but not real. Since $AA^{-1} = I_n$, we have

$$\begin{aligned} 1 &= \sum_{k=1}^{n} a_{ik}a_{ki}^{-1} \\ &= a_{i1}a_{1i}^{-1} + \sum_{k=2}^{n} a_{ik}a_{ki}^{-1}. \end{aligned}$$

The term $a_{i1}a_{1i}^{-1}$ is not real, and so can't be the last summation, in order to cancel out the imaginary components on the right-hand side and sum up to 1. This means that some other element from the ith row of A can't be real, which contradicts our assumption. Using the identity $A^{-1}A = I_n$ and a similar reasoning we can show that

all elements a_{1j}, where $2 \leq j \leq n$, are real. This fact implies, that in the following equation

$$1 = a_{11}a_{11}^{-1} + \sum_{k=2}^{n} a_{1k}a_{k1}^{-1}$$

the last summation is a real number, so a_{11} has to be real as well.

D.2-6 *The errata clarifies, that in the second part of the exercise we also need to assume that A is a symmetric $n \times n$ matrix.*

We have

$$\begin{aligned}(A^{-1})^{\mathrm{T}} &= (A^{\mathrm{T}})^{-1} \\ &= A^{-1},\end{aligned}$$

so A^{-1} is symmetric.

In the second part, since A is symmetric,

$$\begin{aligned}(BAB^{\mathrm{T}})^{\mathrm{T}} &= (B^{\mathrm{T}})^{\mathrm{T}}A^{\mathrm{T}}B^{\mathrm{T}} \\ &= BAB^{\mathrm{T}},\end{aligned}$$

so the product BAB^{T} gives a symmetric matrix.

D.2-7 Let A be an $m \times n$ matrix that has full column rank and let x be an n-vector such that $Ax = 0$. If we denote the columns of A by $a_1, a_2, \ldots, a_n$, then the product Ax can be expressed as the linear combination $\sum_{j=1}^{n} a_j x_j$. Because the column rank of the matrix A is n, all the column vectors are linearly independent, so x_j — viewed as coefficients from the definition of linear dependency — are all 0. Therefore, $x = 0$.

For the other direction, assume that for an $m \times n$ matrix A, the condition $Ax = 0$ implies $x = 0$, and that the matrix A has the rank less than n. This means that there is some set of column vectors of A which are linearly dependent. Note that we can extend this set to include all column vectors of A, $a_1, a_2, \ldots, a_n$, and they all will remain linearly dependent. Let $c_1, c_2, \ldots, c_n$ be coefficients, not all of which are 0, such that $\sum_{j=1}^{n} a_j c_j = 0$. If we consider the n-vector $c = (c_j)$, then the equation can be viewed as $Ac = 0$. But $c \neq 0$, so we have a contradiction.

D.2-8 Let A be an $m \times p$ matrix and let B be $p \times n$ matrix. From the alternate definition of the rank, $r = \mathrm{rank}(AB)$ is the smallest number such that $AB = CD$ for some matrices C and D of respective sizes $m \times r$ and $r \times n$. Let $r' = \mathrm{rank}(A)$ and $r'' = \mathrm{rank}(B)$ and let C', D', C'', D'' be matrices of respective sizes $m \times r'$, $r' \times p$, $p \times r''$, $r'' \times n$ such that $A = C'D'$ and $B = C''D''$. Since $AB = (C'D'C'')D''$ and $C'D'C''$

is an $m \times p$ matrix, we have that $r \leq r'$. Similarly, if we represent the product as $AB = C'(D'C''D'')$, we conclude that $r \leq r''$. Therefore, $r \leq \min\{r', r''\}$.

Now consider a special case, where $p = m$ and A is nonsingular. Let r, r', r'', C and D have the same meaning as before. Of course, $r'' \leq \min\{m, n\} \leq m$ and, by Theorem D.1, $r' = m$. Also,

$$\begin{aligned} B &= A^{-1}AB \\ &= (A^{-1}C)D, \end{aligned}$$

and since $A^{-1}C$ is an $m \times r$ matrix, we have that $r'' \leq r$. Combining this with the inequality $r \leq r''$ shown in the proof for a general case, we obtain $r = r'' = \min\{r', r''\}$. The proof for the case in which $p = n$ and B is nonsingular, is symmetric.

Problems

D-1 *Vandermonde matrix*

The proof is by induction on n and is based on the fact that if we add to a column of a matrix the product by a scalar of another column, then the determinant remains unchanged. This fact can be devised from Theorem D.4 — first multiply a column by a scalar, then add this column to another column, then divide the original column by the scalar.

For the base case of the induction, let $n = 1$. Then the single entry of $V(x_0)$ is 1. Clearly, $\det(V(x_0)) = 1$, that matches the empty product $\prod_{0 \leq j < k \leq 0}(x_k - x_j)$, by convention equal to 1.

For the inductive step, let $n \geq 2$. As the hint suggests, for $i = n-1, n-2, \ldots, 1$, we will multiply column i by $-x_0$ and add it to column $i+1$, to obtain the matrix W shown below:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & x_1 - x_0 & x_1(x_1 - x_0) & x_1^2(x_1 - x_0) & \cdots & x_1^{n-2}(x_1 - x_0) \\ 1 & x_2 - x_0 & x_2(x_2 - x_0) & x_2^2(x_2 - x_0) & \cdots & x_2^{n-2}(x_2 - x_0) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n-1} - x_0 & x_{n-1}(x_{n-1} - x_0) & x_{n-1}^2(x_{n-1} - x_0) & \cdots & x_{n-1}^{n-2}(x_{n-1} - x_0) \end{pmatrix}.$$

By the definition of determinant, $\det(W) = \det(W_{[11]})$. As all the entries in the ith row of $W_{[11]}$ have a factor of $x_i - x_0$, using Theorem D.4 we can take these factors out

and obtain

$$\begin{aligned}\det(V(x_1, x_2, \ldots, x_{n-1})) &= \det(W) \\ &= \det(W_{[11]}) \\ &= \prod_{i=1}^{n-1}(x_i - x_0) \cdot \det(V(x_1, x_2, \ldots, x_{n-1})) \\ &= \prod_{i=1}^{n-1}(x_i - x_0) \prod_{1 \le j < k \le n-1}(x_k - x_j) \\ &= \prod_{0 \le j < k \le n-1}(x_k - x_j).\end{aligned}$$

D-2 ***Permutations defined by matrix-vector multiplication over GF(2)***

a. Let $a_0, a_1, \ldots, a_{n-1}$ be the columns of A and let $J \subseteq \{0, 1, \ldots, n-1\}$ be the set of indices of all linearly independent columns of A. Consider vectors $x, x' \in S_n$, such that there exists $j \in J$ for which $x_j \neq x'_j$, and that for all $j \in \{0, 1, \ldots, n-1\} - J$, $x_j = x'_j = 0$. Then,

$$\begin{aligned}Ax - Ax' &= A(x - x') \\ &= \sum_{j=0}^{n-1} a_j (x_j - x'_j) \\ &= \sum_{j \in J} a_i (x_j - x'_j) \\ &\neq 0,\end{aligned}$$

so $Ax \neq Ax'$. Because $|J| = r$, there are 2^r ways of picking the vector x, and each such x is mapped to a distinct value Ax. Therefore, $|R(A)| \geq 2^r$.

To get the upper bound, observe that any column a_j, where $j \notin J$, is in fact a linear combination of the r linearly independent columns of A. Thus, for any vector x,

$$\begin{aligned}Ax &= \sum_{j=0}^{n-1} a_j x_j \\ &= \sum_{j \in J} a_j x'_j\end{aligned}$$

for some coefficients x'_j. In other words, for any x there exists a vector x', with zeros on positions $j \notin J$, mapped to the same value as x. There are 2^r ways of choosing

such x', so $|R(A)| \leq 2^r$.

If $\text{rank}(A) < n$, then $|R(A)| < 2^n = |S_n|$, so A can't define a permutation on S_n.

b. We showed in part (a), that the value produced by multiplying a vector x by A does not depend on the $n-r$ entries of x occupying the positions corresponding to the linear dependent columns of A. Therefore, for any $y \in R(A)$, $|P(A, y)| \geq 2^{n-r}$.

On the other hand, there are 2^r elements in $R(A)$, each with the preimage consisting of at least 2^{n-r} elements. This means, there are at least $2^r \cdot 2^{n-r} = 2^n$ elements in all preimages in total. Since $|S_n| = 2^n$, each preimage must have exactly 2^{n-r} elements.

c. First observe that $B(S', m)$ consists of size-2^m blocks which contain a value produced by any $x \in S$. Since the first m rows of A only affect the first m entries of Ax, they can affect the value of Ax by at most $2^m - 1$. The block where Ax ends up is therefore determined by the last $n - m$ rows of A. What is more, the numbers in block i differ from the numbers in block 0 by $i2^m$, therefore the produced values are also shifted by a constant offset. Thus, without loss of generality, we may assume that S is block 0, so that only the first m columns of A are relevant during multiplication.

By similar reasoning as in part (a), only the linearly independent rows of the lower left $(n-m)\times m$ submatrix of A can determine the block where Ax will end up, since the entries of Ax on the positions that correspond to the other rows are linear combinations of the other entries of Ax. Thus, Ax spans 2^r blocks, so $|B(S', m)| = 2^r$.

Similarly, we could identify the r linearly independent columns of the submatrix that decide on the block for Ax. A given block is uniquely chosen by a combination of the r entries of x that are multiplied by the elements of these columns, while the remaining $m - r$ entries (besides zeros at positions m, $m + 1$, ..., $n - 1$) can be arbitrary. Thus, each block must be hit by the same number of times, equal to 2^{m-r}.

d. The number of linear permutations is bounded above by the number of pairs (A, c), where A is an $n \times n$ 0-1 matrix and c is an n-bit vector. There are $2^{n^2} \cdot 2^n = 2^{n(n+1)}$ of such pairs. On the other hand, there are $(2^n)!$ permutations of S_n. For $n \geq 3$, $(2^n)^{n+1} \leq (2^n)!$.

e. Observe that $A \cdot 0 + c = c$, and $Ax = \sum_{j=0}^{n-1} a_j x_j$, where $a_0, a_1, \ldots, a_{n-1}$ are the columns of A, so if we let $x = 2^i$, then $Ax = a_i$. Therefore, the linear permutation induced by the pair $(A, 0)$ maps $x = 0$ to 0, and maps $x = 2^i$ to the number whose binary representation is the ith column of A.

Consider a permutation $\pi_{A,c}$ of S_2, such that $\pi_{A,c}(x) = x$, for $x \in \{0, 1, 2\}$. Based on the above observation, this is enough to determine both the matrix A and

the vector c: $A = I$, $c = 0$. These in turn determine the value $\pi_{A,c}(3) = 3$, which makes $\pi_{A,c}$ the identity permutation. Therefore, any permutation of S_2 that maps each $x \in \{0, 1, 2\}$ to x and 3 to anything else than 3, is impossible to achieve by any linear permutation.

Bibliography

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press, 4th edition, 2022.