Probability and Computing : Solutions to Selected Exercises

Michael Mitzenmacher and Eli Upfal

Do Not Distribute!!!

(Current) List of Solved Exercises

Chapter 1 : 1.1 1.2 1.4 1.5 1.6 1.9 1.12 1.13 1.15 1.18 1.21 1.22 1.23 1.25 1.26

Chapter 2 : 2.1 2.5 2.6 2.7 2.8 2.10 2.13 2.14 2.15 2.17 2.18 2.19 2.20 2.21 2.24 2.32

Chapter 3 : 3.2 3.3 3.7 3.10 3.11 3.12 3.16 3.17 3.20 3.25

Chapter 4 : 4.3 4.6 4.7 4.9 4.10 4.11 4.12 4.13 4.19 4.21

Chapter 5 : 5.1 5.4 5.7 5.9 5.11 5.16 5.21 5.22 5.23 [Discussion of Exploratory Assignment]

Chapter 6 : 6.2 6.3 6.4 6.5 6.9 6.10 6.13 6.14 6.16

Chapter 7 : 7.2 7.12 7.17 7.20 7.21 7.23 7.26 7.27

We are still in the process of adding solutions. Expect further updates over time.

Chapter 1

Exercise 1.1: We flip a fair coin ten times. Find the probability of the following events.

(a) The number of heads and the number of tails are equal.

(b) There are more heads than tails.

(c) The ith flip and the $(11 - i)$th flip are the same for $i = 1, \ldots, 5$.

(d) We flip at least four consecutive heads.

Solution to Exercise 1.1:

(a) The probability is $\binom{10}{5}/2^{10} = 252/1024 = 63/256$.

(b) By part (a), the probability that the number of heads and the number of tails is different is $193/256$. By symmettry, the probability that there are more heads than tails is $1/2$ of this, or $193/512$.

(c) The probability that each pair is the same is $1/2$; by independence, the probability that all five pairs are the same is $1/32$.

(d) While there are other ways of solving this problem, with only 1024 possibilities, perhaps the easiest way is just to exhaustively consider all 1024 possibilities (by computer!). This gives $251/1024$.

Exercise 1.2: We roll two standard six-sided dice. Find the probability of the following events, assuming that the outcomes of the rolls are independent.

(a) The two dice show the same number.

(b) The number that appears on the first die is larger than the number on the second.

(c) The sum of the dice is even.

(d) The product of the dice is a perfect square.

Solution to Exercise 1.2:

(a) The probability the second die matches the first is just $1/6$.

(b) By part (a), the probability that the rolls are different is $5/6$. By symmettry, the probability that the first die is larger is $5/12$.

(c) The probability is $1/2$. This can be done by considering all possibilities exhaustively. Alternatively, if the first die comes up with an odd number, the probability is $1/2$ that the second die is also odd and the sum is even. Similarly, if the first die comes up with an even number, the probability is $1/2$ that the second die is also even and the sum is even. Regardless of the outcome of the first roll, the probability is $1/2$.

(d) The possible squares are 1, 4, 9, 16, 25, and 36. There is 1 way for the product to be 1, 9, 16, 25, or 36, and 3 ways for the product to be 4. This gives a total probability of $8/36 = 2/9$.

Exercise 1.4: We are playing a tournament in which we stop as soon as one of us wins n games. We are evenly matched, so each one of us wins any game with probability $1/2$, independently of other games. What is the probability that the loser has won k games when the match is over?

Solution to Exercise 1.4: Suppose that you win, and I have won $k < n$ games. For this to happen, you must win the last game, and in the remaining $n + k - 1$ games, I must win exactly k of them. This probability is just

$$\frac{\binom{n+k-1}{k}}{2^{n+k}}.$$

Of course, I could win, which has the same probability by symmetry. Hence the total desired probability is

$$\frac{\binom{n+k-1}{k}}{2^{n+k-1}}.$$

Exercise 1.5: After lunch one day, Alice suggests to Bob the following method to determine who pays. Alice pulls three six-sided dice from her pocket. These dice are not the standard dice, but have the following numbers on their faces:

- Die A: 1,1,6,6,8,8

- Die B: 2,2,4,4,9,9

- Die C: 3,3,5,5,7,7

The dice are fair, so each side comes up with equal probability. Alice explains that Alice and Bob will each pick up one of the dice. They will each roll their die, and the one who rolls the lowest number loses and will buy lunch. So as to take no advantage, Alice offers Bob the first choice of the dice.

(a) Suppose that Bob chooses Die A and Alice chooses Die B. Write out all of the possible events and their probabilities, and show that the probability that Alice wins is bigger than $1/2$.

(b) Suppose that Bob chooses Die B and Alice chooses Die C. Write out all of the possible events and their probabilities, and show that the probability that Alice wins is bigger than $1/2$.

(c) Since Die A and Die B lead to situations in Alice's favor, it would seem that Bob should choose Die C. Suppose that Bob chooses Die C and Alice chooses Die A. Write out all of the possible events and their probabilities, and show that the probability that Alice wins is still bigger than $1/2$.

Solution to Exercise 1.5: By enumerating all cases, we find the second player wins with probability $5/9$ in all cases. For example, if Bob chooses Die A and Alice chooses Die B, there are nine equally likely outcomes:

$$(1,2),(1,4),(1,9),(6,2),(6,4),(6,9),(8,2),(8,4),(8,9).$$

By counting, we see Alice wins in five of the nine situations.

It might seem odd that Alice always wins. There is a literature on this subject, which is known as non-transitive dice.

Exercise 1.6: Consider the following balls and bin game. We start with 1 black ball and 1 white ball in a bin. We repeatedly do the following: we choose one ball from the bin uniformly at random, and put the ball back in the bin with another ball of the same color. We repeat until there are n balls in the bin. Show that the number of white balls is equally likely to be any number between 1 and $n-1$.

Solution to Exercise 1.6: Let $E_{n,k}$ be the event that there are k white balls when the bin has n balls. We want to show that $Pr[E_{n,k}] = \frac{1}{n-1}$ when $1 \le k < n$ and 0 otherwise. We prove this by induction on n. The claim holds when $n = 2$. Suppose it holds for $n = m$. For $n = m + 1$, condition on the state after m balls. For $E_{m+1,k}$ to occur, either $E_{m,k}$ or $E_{m,k-1}$ occurs. In the first case, we must choose a black ball last (with probability $(m-k)/m$). In the second case, we must choose a white ball last (with probability $(k-1)/m$). By the inductive hypothesis,

$$\begin{aligned} Pr[E_{m+1,k}] &= Pr[E_{m+1,k}|E_{m,k}] \cdot Pr[E_{m,k}] + Pr[E_{m+1,k}|E_{m,k-1}] \cdot Pr[E_{m,k-1}] \\ &= \frac{m-k}{m} \cdot \frac{1}{m-1} + \frac{k-1}{m} \cdot \frac{1}{m-1} = \frac{1}{m}, \end{aligned}$$

as desired. Since $\sum_{k=1}^{m} Pr[E_{m+1,k}] = 1$, $Pr[E_{m+1,k}] = 0$ for all other values of k.

Exercise 1.9: Suppose that a fair coin is flipped n times. For $k > 0$, find an upper bound on the probability that there is a sequence of $\log_2 n + k$ consecutive heads.

3

Solution to Exercise 1.9: Let us assume n is a power of 2. The probability of having $\log_2 n + k$ consecutive heads starting from the jth flip is

$$\frac{1}{2^{\log_2 n + k}} = \frac{1}{2^k n}.$$

Here j can range from 1 to $n - \log_2 n - k + 1$; for $j > n - \log_2 n - k + 1$, there are not enough flips to have k in a row before reaching the nth flip. Using a union bound, we have an upper bound of

$$\frac{n - \log_2 n - k + 1}{2^k n} \leq \frac{1}{2^k}.$$

Exercise 1.12: The following problem is known as the Monty Hall problem, after the host of the game show "Let's Make a Deal." There are three curtains. Behind one curtain is a new car, and behind the other two are goats. The game is played as follows. The contestant chooses the curtain that she thinks the car is behind. Monty then opens one of the other curtains to show a goat. (Monty may have more than one goat to choose from; in this case, assume he chooses which goat to show uniformly at random.) The contestant can then stay with the curtain she originally chose or switch to the other unopened curtain. After that, the location of the car is revealed, and the contestant wins the car or the remaining goat. Should the contestant switch curtains or not, or does it make no difference?

Solution to Exercise 1.12:

We assume that the car is behing a curtain chosen uniformly at random. Since the contestant can pick from 1 car and 2 goats, the probability of choosing a curtain with a car behind it is $1/3$ and that of choosing a curtain with a goat behind it is $2/3$; if the contestant doesn't switch, the probability of winning is just $1/3$.

Now notice that if the contestant switches, she will win whenever she started out by picking a goat! It follows that switching wins $2/3$ of the time, and switching is a much better strategy.

This can be also set up as a conditional probability question, but the above argument is perhaps the simplest; or simply playing the game several times (using cards – an Ace for the car, deuces for the goats) can be very convincing.

Exercise 1.13: A medical company touts its new test for a certain genetic disorder. The false negative rate is small: if you have the disorder, the probability that the test returns a positive result is 0.999. The false positive rate is also small: if you do not have the disorder, the probability that the test returns a positive result is only 0.005. Assume that 2 percent of the population has the disorder. If a person chosen uniformly from the population is tested, and the result comes back positive, what is the probability that the person has the disorder?

Solution to Exercise 1.13: Let X be the event that the person has the disorder, and Y be the event that the test result is positive. Then

$$\Pr(X \mid Y) = \frac{\Pr(X \cap Y)}{\Pr(Y)} = \frac{\Pr(X \cap Y)}{\Pr(X \cap Y) + \Pr(\bar{X} \cap Y)}.$$

We now plug in the appropriate numbers:

$$\Pr(X \mid Y) = \frac{0.02 \cdot 0.999}{0.02 \cdot 0.999 + 0.98 \cdot 0.005} = \frac{999}{1244} \approx 0.803.$$

Exercise 1.15: Suppose that we roll ten standard six-sided dice. What is the probability that their sum will be divisible by 6, assuming that the rolls are independent? (*Hint*: use the principle of deferred decisions, and consider the situation after rolling all but one of the dice.)

Solution to Exercise 1.15: The answer is $1/6$. Using the principle of deferred decisions, consider the situation after the first 9 die rolls. The remainder when dividing the sum of the nine rolls by 6 takes on one of the values $0, 1, 2, 3, 4,$ or 5. If the remainder is 0, the last roll needs to be a 6 to have the final sum be divisible by 6; if the remainder is 1, the last roll needs to be a 5 to have the final sum be divisible by 6; and so on. For any value of the remainder, there is exactly one roll which will make the final sum divisible by 6, so the probability is $1/6$. (If desired, this can be made more formal using conditional probability.)

Exercise 1.18: We have a function $F : \{0, \ldots, n-1\} \to \{0, \ldots, m-1\}$. We know that for $0 \leq x, y \leq n-1$, $F((x + y) \bmod n) = (F(x) + F(y)) \bmod m$. The only way we have to evaluate F is to use a look-up table

that stores the values of F. Unfortunately, an Evil Adversary has changed the value of 1/5 of the table entries when we were not looking.

Describe a simple randomized algorithm that, given an input z, outputs a value that equals $F(z)$ with probability at least 1/2. Your algorithm should work for every value of z, regardless of what values the Adversary changed. Your algorithm should use as few lookups and as little computation as possible.

Suppose I allow you to repeat your initial algorithm three times. What should you do in this case, and what is the probability that your enhanced algorithm returns the correct answer?

Solution to Exercise 1.18: Choose x randomly and let $y = z - x \bmod n$. Look up $F(x)$ and $F(y)$ and output $(F(x) + F(y)) \bmod n$. By the union bound, the probability of an error is bounded by the sum of the probability that $F(x)$ was changed and the probability that $F(y)$ was changed. This sum is 2/5, so $(F(x) + F(y)0 + F(z) \bmod n$ with probability at least 3/5. The algorithm uses two lookups and one random choice. Notice that the probability that $F(x)$ and $F(y)$ are both correct is not $(4/5)^2$; this calculation assumes independence. Only x is chosen randomly, while y is computed from x, so (x, y) is not a random pair.

Suppose we run the algorithm three times independently and return the majority answer (if there is one). This scheme returns the correct answer if at least two of the trials returns the correct answer. By independence, exactly two trials are correct with probability at least $3 \cdot (\frac{3}{5})^2 \cdot \frac{2}{5} = \frac{54}{125}$ and all three trials are correct with probability at least $(\frac{3}{5})^3 = \frac{27}{125}$, so the probability of a correct answer is at least $\frac{81}{125}$.

Exercise 1.21: Given an example of three random events X, Y, Z for which any pair is independent but all three are not mutually independent.

Solution to Exercise 1.21: The standard example is to let X and Y be independent random bits (0 with probability 1/2, 1 with probability 1/2) and let Z be their exclusive-or (equivalently, their sum modulo 2). Independence follows from calculations such as

$$\Pr(Z = 0 \mid X = 0) = \Pr(Y = 0) = 1/2,$$

so X is independent from Z, and similarly Y is. The three values are clearly not independent since Z is determined by X and Y. This construction generalizes; if Z is the exclusive-or of independent random bits $X_1, \ldots, X_n$, then any collection of n of the variables are independent, but the $n + 1$ variables are not.

Exercise 1.22:

(a) Consider the set $\{1, \ldots, n\}$. We generate a subset X of this set as follows: a fair coin is flipped independently for each element of the set; if the coin lands heads, the element is added to X, and otherwise it is not. Argue that the resulting set X is equally likely to be any one of the 2^n possible subsets.

(b) Suppose that two sets X and Y are chosen independently and uniformly at random from all the 2^n subsets of $\{1, \ldots, n\}$. Determine $\Pr(X \subseteq Y)$ and $\Pr(X \cup Y = \{1, \ldots, n\})$. (Hint: use the first part of this problem.)

Solution to Exercise 1.22: For the first part, there are 2^n possible outcomes for the n flips, and each gives a different set; since there are 2^n sets, each set must come up with probability $1/2^n$.

For the second part, suppose we choose sets X and Y using the coin flipping method given in the first part. In order for $X \subseteq Y$, there must be no element that is in X but not in Y. For each of the n elements, the probability that is in X but not in Y is 1/4, and these events are independent for each element. Hence the probability that $X \subseteq Y$ is $(1 - 1/4)^n = (3/4)^n$. Similarly, the probability that $X \cup Y$ contains all elements is $(3/4)^n$, since the probability that each element is in X or Y (or both) is 3/4 independently for each item.

Exercise 1.23: There may be several different min-cut sets in a graph. Using the analysis of the randomized min-cut algorithm, argue that there can be at most $n(n-1)/2$ distinct min-cut sets.

Solution to Exercise 1.23: We have found that the probability that any specific min-cut set is returned is at least $2/(n(n-1))$. If there are k distinct min-cut sets, the probability that any of them is returned is $2k/(n(n-1))$, and we must have

$$\frac{2k}{n(n-1)} \le 1.$$

We can conclude that there are at most $k \le n(n-1)/2$ distinct min-cut sets.

Exercise 1.25 To improve the probability of success of the randomized min-cut algorithm, it can be run multiple times.

(a) Consider running the algorithm twice. Determine the number of edge contractions and bound the probability of finding a min-cut.

(b) Consider the following variation. Starting with a graph with n vertices, first contract the graph down to k vertices using the randomized min-cut algorithm. Make copies of the graph with k vertices, and now run the randomized algorithm on this reduced graph ℓ times, independently. Determine the number of edge contractions and bound the probability of finding a minimum cut.

(c) Find optimal (or at least near-optimal) values of k and ℓ for the variation above that maximize the probability of finding a minimum cut while using the same number of edge contractions as running the original algorithm twice.

Solution to Exercise 1.25

(a) There are $2n - 4$ edge contractions. The probability of success if each run is independent is at least:

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

(b) First consider the probability that no edge from min-cut is contracted in the first $n - k$ contraction steps. In the notation of the book, this is

$$
\begin{aligned}
\Pr(F_{n-k}) &= \Pr(E_{n-k} \mid F_{n-k-1}) \cdot \Pr(E_{n-k-1} \mid F_{n-k-2}) \cdots \Pr(E_2 \mid F_1) \Pr(F_1) \\
&\ge \prod_{i=1}^{n-k} \frac{n-i-1}{n-i+1} \\
&= \frac{k(k-1)}{n(n-1)}.
\end{aligned}
$$

Now consider the last $k - 2$ contractions. The probability that no min-cut edge is contracted in these steps is:

$$
\begin{aligned}
\Pr(F_{n-2} \mid F_{n-k}) &= \Pr(E_{n-2} \mid F_{n-3}) \cdot \Pr(E_{n-3} \mid F_{n-4}) \cdots \Pr(E_{n-k+1} \mid F_{n-k}) \\
&\ge \prod_{i=n-k+1}^{n-2} \frac{n-i-1}{n-i+1} \\
&= \frac{2}{k(k-1)}.
\end{aligned}
$$

Alternatively, this follows by thinking of the subproblem as a case of running the original min-cut algorithm on k vertices.

It follows that if we repeat the last $k - 2$ contractions independent ℓ times the probability of success is at least

$$\Pr(F_{n-k}) \cdot \left(1 - (1 - \Pr(F_{n-2} \mid F_{n-k}))^\ell\right) \ge \frac{k(k-1)}{n(n-1)} \left(1 - \left(1 - \frac{2}{k(k-1)}\right)^\ell\right).$$

(c) Assuming that k and n are large enough so that we can ignore constants, the expression from the previous part is approximately

$$\frac{k^2}{n^2} \left(1 - \left(1 - \frac{2}{k^2}\right)^\ell\right),$$

and we require $2n - 4 = (n - k) + \ell(k - 2)$, or roughly speaking $k\ell \approx n$. We note that when $k = c_1 n^{1/3}$ and $\ell = c_2 n^{2/3}$ the resulting probability of success is at least $\Omega(n^{-4/3})$, and this is essentially to the best possible. This can be checked by various programs like Mathematica. Alternatively, note that when $\ell < k^2$, as an approximation, we hvae

$$\left(1 - \left(1 - \frac{2}{k^2}\right)^{\ell}\right) \approx \frac{2\ell}{k^2},$$

and hence the success probability is at least (approximately) $2\ell/n^2$, suggesting ℓ should be as large as possible. However, once ℓ is large enough so that $(1 - (1 - 2/k^2)^{\ell})$ is constant, increasing ℓ can only improve this term by at most a constant factor, suggesting k should be as large as possible. This gives the tradeoff giving the result.

Exercise 1.26: Tic-tac-toe always ends up in a tie if players play optimally. Instead, we may consider random variations of tic-tac-toe.

(a) First variation: Each of the nine squares is labeled either X or O according to an independent and uniform coin flip. If only one of the players has one (or more) winning tic-tac-toe combinations, that player wins. Otherwise, the game is a tie. Determine the probability that X wins. (You may want to use a computer program to help run through the configurations.)

(b) Second variation: X and O take turns, with the X player going first. On the X player's turn, an X is placed on a square chosen independently and uniformly at random from the squares that are still vacant; O plays similarly. The first player to have a winning tic-tac-toe combination wins the game, and a tie occurs if neither player achieves a winning combination. Find the probability that each player wins. (Again, you may want to write a program to help you.)

Solution to Exercise 1.26: For the first variation, out of $2^9 = 512$ possible games, there are 116 ties. By symmetry, X and O win with equal probability, so each wins 198 of the possible games, giving a winning probability of roughly .386.

For the second variation, there are 255168 possible games. X wins 131184; Y wins 77904, leaving 46080 ties. However, shorter games occur with higher probability. Specifically, a game with k moves occurs with probability $\frac{1}{9 \cdot 8 \cdots \cdot (9 - k + 1)}$. Taking this into account, X wins with probability roughly 0.585, and Y wins with probability roughly 0.288.

Chapter 2

Exercise 2.1: Suppose we roll a fair k-sided die with the numbers 1 through k on the die's faces. If X is the number that appears, what is $\mathbf{E}[X]$?

Solution to Exercise 2.1:
$$\mathbf{E}[X] = \sum_{i=1}^{k} \frac{1}{k} \cdot i = \frac{k(k+1)}{2k} = \frac{k+1}{2}.$$

Exercise 2.5: If X is a $B(n, 1/2)$ random variable with $n \geq 1$, show that the probability that X is even is $1/2$.

Solution to Exercise 2.5: This can be proven by induction. It is trivially true for $n = 1$. Now note that a $B(n, 1/2)$ random variable corresponds to the number of heads in n fair coin flips. Consider n flips. By the induction hypothesis, the number Y of heads after the first $n-1$ of these coin flips, which has distribution $B(n-1, 1/2)$, is even with probability $1/2$. Flipping one more coin, we find the number of heads is

$$(1/2) \cdot \Pr(Y \text{ is even}) + (1/2) \cdot \Pr(Y \text{ is odd}) = 1/2.$$

In fact, notice we don't even need the induction! Even if just the *last* flip is fair, the same equation holds, and since $\Pr(Y \text{ is even}) + \Pr(Y \text{ is odd}) = 1$ the result holds.

Exercise 2.6: Suppose that we independently roll two standard six-sided dice. Let X_1 be the number that shows on the first die, X_2 the number on the second die, and X the sum of the numbers on the two dice.

(a) What is $\mathbf{E}[X \mid X_1 \text{ is even}]$?

(b) What is $\mathbf{E}[X \mid X_1 = X_2]$?

(c) What is $\mathbf{E}[X_1 \mid X = 9]$?

(d) What is $\mathbf{E}[X_1 - X_2 \mid X = k]$ for k in the range $[2, 12]$?

Solution to Exercise 2.6:

(a) What is $\mathbf{E}[X \mid X_1 \text{ is even}]$?

$$
\begin{aligned}
\mathbf{E}[X \mid X_1 \text{ is even}] &= \sum_{k=2}^{12} \Pr(X = k \mid X_1 \text{ is even}) \\
&= 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{18} + 5 \cdot \frac{2}{18} + \ldots + 12 \frac{1}{18} \\
&= 7.5.
\end{aligned}
$$

There are other ways of calculating this; for example

$$
\begin{aligned}
\mathbf{E}[X \mid X_1 \text{ is even}] &= \mathbf{E}[X_1 + X_2 \mid X_1 \text{ is even}] \\
&= \mathbf{E}[X_1 \mid X_1 \text{ is even}] + \mathbf{E}[X_2 \mid X_1 \text{ is even}] \\
&= 4 + 3.5.
\end{aligned}
$$

(b) What is $\mathbf{E}[X \mid X_1 = X_2]$?
$$\sum_{i=1}^{6} 2i \cdot \frac{1}{6} = 7.$$

(c) What is $\mathbf{E}[X_1 \mid X = 9]$?

There are only 4 combinations giving $X = 9$, all equally likely; over these 4 combinations, X_1 takes on the values from 3 to 6. Hence

$$\mathbf{E}[X_1 \mid X = 9] = \sum_{k} \Pr(X_1 = k \mid X = 9) = \frac{1}{4}(3 + 4 + 5 + 6) = 4.5.$$

(d) What is $\mathbf{E}[X_1 - X_2 \mid X = k]$ for k in the range $[2, 12]$?

Note that $\mathbf{E}[X_1 - X_2 \mid X = k] = \mathbf{E}[X_1 \mid X = k] - \mathbf{E}[X_2 \mid X = k]$. But this difference is 0 by symmetry.

Exercise 2.7: Let X and Y be independent geometric random variables, where X has parameter p and Y has parameter q.

(a) What is the probability that $X = Y$?

(b) What is $\mathbf{E}[\max(X, Y)]$?

(c) What is $\Pr(\min(X, Y) = k)$?

(d) What is $\mathbf{E}[X \mid X \leq Y]$?

You may find it helpful to keep in mind the memoryless property of geometric random variables.

Solution to Exercise 2.7:

(a) Think of flipping biased coins to determine X and Y (according to the frist head for each). The probability that X and Y both equal 1 is pq. The probability that $X = 1$ and $Y > 1$ is $p(1 - q)$, and the probability that $X > 1$ and $Y = 1$ is $q(1 - p)$. Hence if the outcome is decided by the first flips of the coins, we have

$$\Pr(X = Y) = pq/(pq + (1 - p)q + (1 - q)p) = pq/(1 - (1 - p)(1 - q)).$$

If $X > 1$ and $Y > 1$, then by the memoryless property, the distributions of X and Y are as if we are just beginning again. Hence, whenever we stop, we must have

$$\Pr(X = Y) = pq/(pq + (1 - p)q + (1 - q)p) = pq/(1 - (1 - p)(1 - q)).$$

(b) Think of flipping biased coins to determine X and Y. Let $Z = \max(X, Y)$. Consider the four cases for the outcomes from the first flip of X and Y. For example, if $X > 1$ and $Y > 1$, then by the memoryless property the remaining distribution for the number of flips until the first head for each coin remains geometric. If $X = 1$ and $Y > 1$ then we only consider the number of flips until the first head until the second coin. Following this logic for all four cases, we have:

$$
\begin{aligned}
\mathbf{E}[Z] &= (pq) \cdot 1 + p(1 - q)(1 + \mathbf{E}[Y]) + (1 - p)q(1 + \mathbf{E}[X]) + (1 - p)(1 - q)(1 + \mathbf{E}[Z]) \\
&= 1 + p(1 - q)/q + q(1 - p)/p + (1 - p)(1 - q)\mathbf{E}[Z].
\end{aligned}
$$

This gives
$$\mathbf{E}[Z] = (1 + p(1 - q)/q + q(1 - p)/p)/(1 - (1 - p)(1 - q)).$$

(c) What is $\Pr(\min(X, Y) = k)$? Think of flipping biased coins to determine X and Y. On the first flip, at least one of the two coins is heads with probability $s = 1 - (1 - p)(1 - q)$. If not, by the memoryless property, the distribution of flips until the first head for each coin remains geometric. It follows that $\min(X, Y)$ is itself geometric with parameter s, giving the distribution.

(d) Using the now familair logic of considering whether $Y \geq X$ is determined by the first flip of respective coins, we find that

$$\Pr(Y \geq X) = (pq + p(1 - q))/(pq + (1 - p)q + (1 - q)p) = p/(1 - (1 - p)(1 - q)).$$

The expectation is then

$$
\begin{aligned}
\mathbf{E}[X \mid X \leq Y] &= \sum_{x=1}^{\infty} x \Pr(X = x \mid Y \geq X) \\
&= \sum_{x=1}^{\infty} x \Pr(Y \geq X = x) / \Pr(Y \geq X) \\
&= \sum_{x=1}^{\infty} x \frac{(1-p)^{x-1}p(1-q)^{x-1}}{p/(1-(1-p)(1-q))} \\
&= (1-(1-p)(1-q)) \sum_{x=0}^{\infty} x((1-p)(1-q))^{x-1} \\
&= \frac{1}{1-(1-p)(1-q)}.
\end{aligned}
$$

Exercise 2.8:

(a) Alice and Bob decide to have children until either they have their first girl or they have $k \geq 1$ children. Assume that each child is a boy or girl independently with probability $1/2$, and that there are no multiple births. What is the expected number of female children that they have? What is the expected number of male children that they have?

(b) Suppose Alice and Bob simply decide to keep having children until they have their first girl. Assuming that this is possible, what is the expected number of boy children that they have?

Solution to Exercise 2.8:

(a) With probability $1/2^k$, they have 0 girls; with probability $1-1/2^k$, they have 1 girl. Hence the expected number of female children is $1 - 1/2^k$. The probability that they have at least $i \geq 1$ boys is $1/2^i$, since this occurs whenever the first i children are boys. Hence the expected number of male children is

$$
\sum_{i=1}^{k} \frac{1}{2^i} = 1 - 1/2^k.
$$

The two values match.

(b) Extending the argument from the previous part, the expected number of boys is

$$
\sum_{i=0}^{\infty} \frac{1}{2^{i+1}} = 1.
$$

Hence, even if they have kids until their first girl, on average they'll have just 1 boy.

Exercise 2.10:

(a) Show by induction that if $f : \mathbf{R} \to \mathbf{R}$ is convex then, for any $x_1, x_2, \ldots, x_n$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$ with $\sum_{i=1}^{n} \lambda_i = 1$,

$$
f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).
$$

(b) Use Eqn. (2.2) to prove that if $f : \mathbf{R} \to \mathbf{R}$ is convex then

$$
\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])
$$

for any random vairable X that takes on only finitely many values.

Solution to Exercise 2.10:

(a) This is done by induction. The case $n = 1$ is trivial; the case $n = 2$ is just the definition of convexity.

Assume that the statement holds for some value n. Consider now $x_1, x_2, \ldots, x_n, x_{n+1}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n, \lambda_{n+1}$ with $\sum_{i=1}^{n+1} \lambda_i = 1$, If $\lambda_{n+1} = 1$ the formula holds trivially, so we many assume $\lambda_{n+1} < 1$. Let $\gamma_i = \lambda_1/(1 - \lambda_{n+1})$. Now

$$
\begin{aligned}
f\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) &= f\left(\sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} x_{n+1}\right) \\
&= f\left((1 - \lambda_{n+1}) \sum_{i=1}^{n} \gamma_i x_i + \lambda_{n+1} x_{n+1}\right) \\
&\leq (1 - \lambda_{n+1}) f\left(\sum_{i=1}^{n} \gamma_i x_i\right) + \lambda_{n+1} f(x_{n+1}).
\end{aligned}
$$

The last line comes from convexity of f (the case of $n = 2$ items). But now by the induction hypothesis

$$
\begin{aligned}
(1 - \lambda_{n+1}) f\left(\sum_{i=1}^{n} \gamma_i x_i\right) + \lambda_{n+1} f(x_{n+1}) &\leq (1 - \lambda_{n+1}) \gamma_i \sum_{i=1}^{n} f(x_i) + \lambda_{n+1} f(x_{n+1}) \\
&= \sum_{i=1}^{n} \lambda_i f(x_i) + \lambda_{n+1} f(x_{n+1}) \\
&= \sum_{i=1}^{n+1} \lambda_i f(x_i),
\end{aligned}
$$

completing the induction.

(b) Let $x_1, \ldots, x_n$ be the set of values that X takes on and let $\lambda_i = \Pr(X = x_i)$. By part (a) we have

$$
\mathbf{E}[f(X)] = \sum_{i=1}^{n} f(x_i) \Pr(X = x_i) \geq f\left(\sum_{i=1}^{n} x_i \Pr(X = x_i)\right) = f(\mathbf{E}[X]).
$$

Exercise 2.13:

(a) Consider the following variation of the coupon collector's problem. Each box of cereal contains one of $2n$ different coupons. The coupons are organized into n pairs, so that coupons 1 and 2 are a pair, coupons 3 and 4 are a pair, and so on. Once you obtain one coupon from every pair, you can obtain a prize. Assuming that the coupon in each box is chosen independently and uniformly at random from the $2n$ possibilities, what is the expected number of boxes you have to buy before you can claim the prize?

(b) Generalize the result of the problem above for the case where there are kn different coupons, organized into n disjoint sets of k coupons, so that you need one coupon from every set.

Solution to Exercise 2.13:

(a) Let X_i be the number of boxes bought while having a coupon from $i - 1$ distinct pairs to obtain a coupon from an ith distinct pair. Then $X = \sum_{i=1}^{n} X_i$ is the desired number of boxes. Each X_i is a geometric random random variable, with $p_i = 1 - (2(i - 1))/2n = 1 - (i - 1)/n$. Hence this has the same behavior as the coupon collector's problem, and the same expectation, $nH(n)$.

(b) Nothing changes; X_i is now the number of boxed needed to go from $i - 1$ groups to i groups, and again $p_i = 1 - (2(i - 1))/2n = 1 - (i - 1)/n$.

Exercise 2.14: The geometric distribution arises as the distribution of the number of times we flip a coin until it comes up heads. Consider now the distribution of the number of flips X until the kth head appears, where each coin flip comes up heads independently with probability p. Prove that this distribution is given by

$$\Pr(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

for $n \geq k$. (This is known as the *negative binomial* distribution.)

Solution to Exercise 2.14: In order for $X = n$, the nth flip must be heads, and from the other $n-1$ flips, we must choose exactly $k-1$ of them to be heads. Hence

$$\Pr(X = n) = p \cdot \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}.$$

Exercise 2.15: For a coin that comes up heads independently with probability p on each flip, what is the expected number of flips until the k-th head?

Solution to Exercise 2.15: Let X_i be the number of flips between the $(i-1)$st and and ith heads. Then each X_i is geometric with expectation $1/p$. Let $X = \sum_{i=1}^{k} X_i$ be the number of flips until the kth head. By linearity of expectations, $\mathbf{E}[X] = \sum_{i=1}^{k} \mathbf{E}[X_i] = k/p$. Note that this gives the expectation of the negative binomial distribution of Exercise 2.14. (The expectation of the negative binomial could also be computed directly, but it is more work.)

Exercise 2.17: Recall the recursive spawning process described in Section 2.3. Suppose that each call to process $\mathcal{S}$ recursively spawns new copies of the process $\mathcal{S}$, where the number of new copies is 2 with probability p and 0 with probability $1 - p$. If Y_i denotes the number of copies of $\mathcal{S}$ in the i-th generation, determine $\mathbf{E}[Y_i]$. For what values of p is the expected total number of copies bounded?

Solution Exercise 2.17: Let Y_i be the number of processes in generation i, and similarly for Y_{i-1}. Let Z_k be the number of processes spawned by the kth process in generation Y_{i-1}. We have

$$
\begin{aligned}
\mathbf{E}[Y_i \mid Y_{i-1} = y_{i-1}] &= \mathbf{E}\left[\sum_{k=1}^{y_{i-1}} Z_k\right] \\
&= \sum_{k=1}^{y_{i-1}} \mathbf{E}[Z_k] \\
&= 2y_{i-1}p.
\end{aligned}
$$

Further,

$$\mathbf{E}[Y_i] = \mathbf{E}[\mathbf{E}[Y_i \mid Y_{i-1}]] = \mathbf{E}[2Y_{i-1}p] = 2p\mathbf{E}[Y_{i-1}].$$

Inductively, since $Y_0 = 1$, we have $\mathbf{E}[Y_i] = (2p)^i$. The expected total number of processes spawned is

$$\mathbf{E}\left[\sum_{i=0}^{\infty} Y_i\right] = \sum_{i=0}^{\infty} \mathbf{E}[Y_i] = \sum_{i=0}^{\infty} (2p)^i,$$

which is finite when $2p < 1$ (or $p < 1/2$).

Exercise 2.18: The following approach is often called *reservoir sampling*. Suppose we have a sequence of items passing by one at a time. We want to maintain a sample of one item that has the property that it is uniformly distributed over all the items that we have seen at each step. Moreover, we want to accomplish this without knowing the total number of items in advance or storing all of the items that we see.

Consider the following algorithm, which stores just one item in memory at all times. When the first item appears, it is stored in the memory. When the kth item appears, it replaces the item in memory with probability $1/k$. Explain why this algorithm solves the problem.

Solution to Exercise 2.18: The proof is by induction. Clearly after the first item appears, the item in memory is uniform over all the items that have appeared. After the kth item appears, with probability

$1 - 1/k = (k-1)/k$ the item in memory is uniform over the first $k-1$ items, by they induction hypothesis. Hence each of the first $k-1$ items is in the memory with probability $1/k$, and by construction so is the kth item, completing the induction.

Exercise 2.19: Suppose that we modify the reservoir sampling algorithm of Exercise 2.18 so that, when the kth item appears, it replaces the item in memory with probability $1/2$. Explain what the distribution of the item in memory looks like.

Solution to Exercise 2.19: By induction, we can again show that the probability that the jth item is in memory after the kth item has appeared is 2^{j-k-1}, except for the first item, which is in memory with probability 2^{-k+1}. (Note $j \geq 1$.)

Exercise 2.20: A permutation on the numbers $[1, n]$ can be represented as a function $\pi : [1, n] \to [1, n]$, where $\pi(i)$ is the position of i in the ordering given by the permutation. A fixed point of a permutation $\pi : [1, n] \to [1, n]$ is a value for which $\pi(x) = x$. Find the expected number of fixed points of a permutation chosen uniformly at random from all permutations.

Solution to Exercise 2.20: Let $X_i = 1$ if i is a fixed point and 0 otherwise. As each element is a fixed point with probability $1/n$ in a random permutation we have $\mathbf{E}[X_i] = 1/n$. Let $X = \sum_{i=1}^{n} X_i$ be the number of fixed points. By linearity of expectations

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = n/n = 1.$$

Exercise 2.21: Let $a_1, a_2, \ldots, a_n$ be a random permutation of $\{1, 2, \ldots, n\}$, equally likely to be any of the $n!$ possible permutations. When sorting the list $a_1, a_2, \ldots, a_n$, the element a_i has to move a distance of $|a_i - i|$ places from its current position to reach its position in the sorted order. Find

$$\mathbf{E}\left[\sum_{i=1}^{n} |a_i - i|\right],$$

the expected total distance elements will have to be moved.

Solution to Exercise 2.21: There are many way to perform the calculation, but the following approach is appealing. Consider a permutation chosen at random, and let $X_{ij} = |j - i|$ if in this permutation $j = a_i$, and 0 otherwise. Then

$$\mathbf{E}\left[\sum_{i=1}^{n} |a_i - i|\right] = \mathbf{E}\left[sum_{1 \leq i,j \leq n} X_{ij}\right] = \sum_{1 \leq i,j \leq n} \mathbf{E}[X_{ij}] = \sum_{1 \leq i,j \leq n} \frac{|j - i|}{n}.$$

The first equality follows from the definition of X_{ij}, the second from the linearity of expectations, and the third from the fact that $X_{ij} = |j - i|$ precisely with probability $1/n$ if the permutation is random. Now to simplify this summmation, note that over the range of i and j values, $|j - i|$ takes on the value 1 exactly $2(n-1)$ times, the value 2 exactly $2(n-2)$ times, and so on; hence the sum equals

$$\sum_{i=1}^{n} \frac{2i(n-i)}{n}.$$

With some algebra, this simplifies to

$$\frac{n^2 - 1}{3}.$$

Exercise 2.24: We roll a standard fair die over and over. What is the expected number of rolls until the first pair of consecutive sixes appears? (Hint: the answer is not 36.)

Solution to Exercise 2.24: One way to think of this problem is the following. Let X be the number of rolls until we see a six and Y be the number of rolls until we see two consecutive sixes. If the next roll following this six is also a six, then we have seen a pair of consecutive sixes. If not, then it is as though we

must start all over again. Hence, with probability $1/6$, $Y = X+1$, and with probability $5/6$, $Y = X+1+Y'$, where Y' has the same distribution as Y. Hence

$$\mathbf{E}[Y] = \frac{1}{6}\mathbf{E}[X+1] + \frac{5}{6}\mathbf{E}[X+1+Y'].$$

Simplifying using linearity of expectations and $\mathbf{E}[Y'] = \mathbf{E}[Y]$ gives

$$\frac{1}{6}\mathbf{E}[Y] = \mathbf{E}[X] + 1.$$

Now $\mathbf{E}[X]$, the expected number of rolls until our first six, equals six. Hence $\mathbf{E}[Y] = 42$.

Chapter 3

Exercise 3.2: Let X be a number chosen uniformly at random from $[-k, k]$. Find $\mathbf{Var}[X]$.

Solution to Exercise 3.2: We have $\mathbf{E}[X] = 0$, so $\mathbf{Var}[X] = \mathbf{E}[X^2]$. We find

$$
\begin{aligned}
\mathbf{Var}[X] &= \sum_{i=-k}^{k} \frac{1}{2k+1} i^2 \\
&= \sum_{i=1}^{k} \frac{2}{2k+1} i^2 \\
&= \frac{2}{2k+1} \frac{k(k+1)(2k+1)}{6} \\
&= \frac{k(k+1)}{3}.
\end{aligned}
$$

Exercise 3.3: Suppose that we roll a standard fair die 100 times. Let X be the sum of the numbers that appear over the 100 rolls. Use Chebyshev's inequality to bound $\Pr(|X - 350| \geq 50)$.

Solution to Exercise 3.3: For a roll with outcome Y, we have $\mathbf{E}[Y] = 7/2$ and

$$
\mathbf{Var}[Y] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \sum_{i=1}^{6} i^2/6 - 49/4 = 35/12.
$$

By linearity of expectations and of variance (when the variables being summed are independent), we have $\mathbf{E}[X] = 350$ and $\mathbf{E}[Y] = 3500/12 = 875/3$. Hence, by Chebyshev's inequality

$$
\Pr(|X - 350| \geq 50) \leq \frac{875}{3 \cdot 2500} = \frac{7}{60}.
$$

Exercise 3.7: A simple model of the stock market suggests that, each day, a stock with price q will increase by a factor $r > 1$ to qr with probability p and will fall to q/r with probability $1 - p$. Assuming we start with a stock with price 1, find a formula for the expected value and the variance of the price of the stock after d days.

Solution to Exercise 3.7: Let X_d be the price of the stock after d days. Let $Y_i = r$ if the price increases on the ith day and let $Y_i = r^{-1}$ otherwise. We assume that the Y_i are independent. Now $X_d = \prod_{i=1}^{d} Y_i$, so

$$
\mathbf{E}[X_d] = \mathbf{E}[\prod_{i=1}^{d} Y_i] = \prod_{i=1}^{d} \mathbf{E}[Y_i] = \left(pr + \frac{1-p}{r} \right)^d,
$$

where we make use the independence to simplify the expectation of the product. Similarly

$$
\mathbf{E}[X_d^2] = \mathbf{E}[\prod_{i=1}^{d} Y_i^2] = \prod_{i=1}^{d} \mathbf{E}[Y_i^2] = \left(pr^2 + \frac{1-p}{r^2} \right)^d.
$$

Hence

$$
\mathbf{Var}[X_d] = \mathbf{E}[X_d^2] - \mathbf{E}[X_d]^2 = \left(pr^2 + \frac{1-p}{r^2} \right)^d - \left(pr + \frac{1-p}{r} \right)^{2d}.
$$

Exercise 3.10: For a geometric random variable X, find $\mathbf{E}[X^3]$ and $\mathbf{E}[X^4]$. (*Hint*: Use Lemma 2.5.)

Solution to Exercise 3.10: Let X be a geometric random variable with parameter p. Let $Y = 1$ if $X = 1$, and $Y = 0$ otherwise. Finally, let $X = X' + 1$. By Lemma 2.5, we have

$$
\begin{aligned}
\mathbf{E}[X^3] &= \Pr(Y = 1)\mathbf{E}[X^3 \mid Y = 1] + \Pr(Y = 0)\mathbf{E}[X^3 \mid Y = 0] \\
&= p \cdot 1 + (1-p)\mathbf{E}[(X'+1)^3 \mid Y = 0].
\end{aligned}
$$

Here we use $\mathbf{E}[X^3 \mid Y = 1] = 1$ since $X = 1$ when $Y = 1$. To simplify the expression further, recall that conditioned on $Y = 0$, or equivalently that $X > 1$, X' has a geometric distribution by the memoryless property. Hence

$$\mathbf{E}[(X' + 1)^3 \mid Y = 0] = \mathbf{E}[(X + 1)^3],$$

and we continue

$$
\begin{aligned}
\mathbf{E}[X^3] &= p \cdot 1 + (1 - p)\mathbf{E}[(X' + 1)^3 \mid Y = 0] \\
&= p + (1 - p)\mathbf{E}[(X + 1)^3] \\
&= p + (1 - p)\mathbf{E}[X^3] + 3(1 - p)\mathbf{E}[X^2] + 3(1 - p)\mathbf{E}[X] + (1 - p) \\
&= (1 - p)\mathbf{E}[X^3] + 1 + 3(1 - p)(2 - p)/p^2 + 3(1 - p)/p.
\end{aligned}
$$

Solving for $\mathbf{E}[X^3]$ gives

$$
\begin{aligned}
\mathbf{E}[X^3] &= 1/p + 3(1 - p)(2 - p)/p^3 + 3(1 - p)/p^2 \\
&= (p^2 - 6p + 6)/p^3.
\end{aligned}
$$

Similarly, we find

$$
\begin{aligned}
\mathbf{E}[X^4] &= \Pr(Y = 1)\mathbf{E}[X^4 \mid Y = 1] + \Pr(Y = 0)\mathbf{E}[X^4 \mid Y = 0] \\
&= (1 - p)\mathbf{E}[X^4] + 1 + 4(1 - p)\mathbf{E}[X^3] + 6(1 - p)\mathbf{E}[X^2] + 4(1 - p)\mathbf{E}[X],
\end{aligned}
$$

giving

$$
\begin{aligned}
\mathbf{E}[X^4] &= 1/p + 4(1 - p)(p^2 - 6p + 6)/p^4 + 6(1 - p)(2 - p)/p^3 + 4(1 - p)/p^2 \\
&= 1/p + 4(1 - p)(p^2 - 6p + 6)/p^4 + 6(1 - p)(2 - p)/p^3 + 4(1 - p)/p^2 \\
&= (24 - 36p + 14p^2 - p^3)/p^4.
\end{aligned}
$$

Exercise 3.11: Recall the Bubblesort algorithm of Exercise 2.22. Determine the variance of the number of inversions that need to be corrected by Bubblesort.

Solution to Exercise 3.11: Denote the input set by $\{1, \ldots, n\}$ and the permutation by π. For $i < j$, let X_{ij} be 1 if i and j are inverted and 0 otherwise. The total number of inversions is given by $X = \sum_{i<j} X_{ij}$, and $\mathbf{E}[X_{ij}]$ is the probability a pair is inverted, which is $1/2$. By linearity of expectations,

$$\mathbf{E}[X] = \sum_{i<j} \mathbf{E}[X_{ij}] = \frac{1}{2}\binom{n}{2}.$$

To compute $\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$, we find

$$\mathbf{E}[X^2] = \mathbf{E}\left[\left(\sum_{i<j} X_{ij}\right)^2\right] = \sum_{i<j} E[X_{ij}^2] + \sum_{\{i,j\}\neq\{k,\ell\}} E[X_{ij}X_{k\ell}],$$

where is the second summation we have $i < j$ and $k < \ell$. Since X_{ij} is an indicator variable, $X_{ij}^2 = X_{ij}$, and the first sum is just $\mathbf{E}[X]$. To compute the second sum, we consider several cases. If none of the four indices are equal, then X_{ij} and $X_{k\ell}$ are independent, so $\mathbf{E}[X_{ij}X_{k\ell}] = \mathbf{E}[X_{ij}]\mathbf{E}[X_{k\ell}] = 1/4$. There are $\binom{n}{2}\binom{n-2}{n}$ such pairs $(i, j), (k, \ell)$.

If the pairs overlap, there are three distinct indices $i < j < k$ and three cases:

(i) $(i, j), (i, k)$: This occurs $2\binom{n}{3}$ ways, and in this case

$$\mathbf{E}[X_{ij}X_{ik}] = \Pr(X_{ij}X_{ik} = 1) = \Pr(\pi(i) > \pi(j) \text{ and } \pi(i) > \pi(k)) = \frac{1}{3}.$$

(ii) $(i, j), (j, k)$: This occurs $2\binom{n}{3}$ ways, and in this case

$$\mathbf{E}[X_{ij}X_{jk}] = \Pr(X_{ij}X_{jk} = 1) = \Pr(\pi(i) > \pi(j) > \pi(k)) = \frac{1}{6}.$$

(iii) $(i, k), (j, k)$: This occurs $2\binom{n}{3}$ ways, and in this case

$$\mathbf{E}[X_{ik}X_{jk}] = \Pr(X_{ik}X_{jk} = 1) = \Pr(\pi(i) > \pi(k) \text{ and } \pi(j) > \pi(k)) = \frac{1}{3}.$$

Therefore

$$\sum_{\{i,j\} \neq \{k,\ell\}} E[X_{ij}X_{k\ell}] = \frac{2}{3}\binom{n}{3} + \frac{1}{3}\binom{n}{3} + \frac{2}{3}\binom{n}{3} = \frac{5}{3}\binom{n}{3}.$$

Finally,

$$Var[X] = \frac{1}{2}\binom{n}{2} + \frac{1}{4}\binom{n}{2}\binom{n-2}{2} + \frac{5}{3}\binom{n}{3} - \frac{1}{4}\binom{n}{2}^2 = \frac{n(n-1)(2n+5)}{72}.$$

Exercise 3.12: Find an example of a random variable with finite expectation and unbounded variance. Give a clear argument showing that your choice has these properties.

Solution to Exercise 3.12: Let X be an integer-valued random variable defined by $\Pr(X = i) = c/i^3$ for $i \geq 1$ and some positive constant c. This is possible since $\sum_{i=1}^{\infty} 1/i^3$ is finite. Then

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} 1/i^2$$

which is finite, but

$$\mathbf{E}[X^2] = \sum_{i=1}^{\infty} 1/i$$

which is not. Hence X has finite expectation but infinite variance.

Exercise 3.16: This problem shows that Markov's inequality is as tight as it could possibly be. Given a positive integer k, describe a random variable X assuming only non-negative values such that

$$\Pr(X \geq k\mathbf{E}[X]) = \frac{1}{k}.$$

Solution to Exercise 3.16: Let $X = 0$ with probability $1 - 1/k$, and let $X = k$ with probability $1/k$. This satisfies the equality.

Looking at the proof of Markov's inequality, one can see that for tightness to hold, the distribution must have this form. That is, X must either be 0 or $k\mathbf{E}[X]$.

Exercise 3.17: Can you give an example (similar to that for Markov's inequality in Exercise 3.16) that shows that Chebyshev's inequality is tight? If not, explain why not.

Solution to Exercise 3.17: In fact you can give such an example. Looking at the proof of Theorem 3.6, tightness will occur whenever

$$\Pr((X - \mathbf{E}[X])^2 \geq a^2) = \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2},$$

which is where Markov's inequality is applied. By the argument for the solution to Exercise 3.16, for Markov's inequality to be tight, we need either $(X - \mathbf{E}[X])^2 = 0$ or $(X - \mathbf{E}[X])^2 = a^2$. Hence for some $m = \mathbf{E}[X]$, we should have that $X = m$, $X = m + a$, or $X = m - a$. In order to have $\mathbf{E}[X] = m$, the distribution should be symmetric, so that $X = m$ with probability p and $X = m - a$ with probability p. Any such distribution would be satisfactory. Students may just come up with the special case where $X = m + a$ with probability $1/2$ and $X = m - a$ with probability $1/2$, which is satisfactory.

Exercise 3.20:

(a) Chebyshev's inequality uses the variance of a random variable to bound its deviation from its expectation. We can also use higher moments. Suppose that we have a random variable X and an even integer k for which $\mathbf{E}[(X - \mathbf{E}[X])^k]$ is finite. Show that

$$\Pr\left(|X - \mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\right) \leq \frac{1}{t^k}.$$

(b) Why is it difficult to derive a similar inequality when k is odd?

Solution to Exercise 3.20: For even k the kth moment is always non-negative, so we have

$$\Pr(|X - \mathbf{E}[X]| > \alpha) = \Pr(|X - \mathbf{E}[X]|^k > \alpha^k) = \Pr((X - \mathbf{E}[X])^k > \alpha^k).$$

This allows also to use Markov inequality:

$$
\begin{aligned}
\Pr(|X - \mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}) &= \Pr((X - \mathbf{E}[X])^k > t^k\mathbf{E}[(X - \mathbf{E}[X])^k]) \\
&\leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^k]}{t^k\mathbf{E}[(X - \mathbf{E}[X])^k]} \\
&= \frac{1}{t^k}.
\end{aligned}
$$

When k is odd, terms like $X - \mathbf{E}[X]$ might be negative, and the argument will not go through. For example, $(X - \mathbf{E}[X])^k$ might be negative and we cannot apply Markov's inequality to it.

Exercise 3.25: The weak law of large numbers says that if $X_1, X_2, X_3, \ldots$ are independent and identically distributed random variables with mean μ and standard deviation σ, then for any constant $\epsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \epsilon\right) = 0.$$

Use Chebyshev's inequality to prove the weak law of large numbers.

Solution to Exercise 3.25:

Let $S_n = X_1 + X_2 + \cdots + X_n$. We have $\mathbf{E}[S_n] = n\mu$, and since the X_i are independent, we have

$$\mathbf{Var}[S_n] = \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + \ldots + \mathbf{Var}[X_n] = n\sigma^2.$$

Fix an $\epsilon > 0$. We have, by Chebyshev's inequality,

$$
\begin{aligned}
\Pr\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) &= \Pr(|S_n - \mu n| > \epsilon n) \\
&\leq \frac{\mathbf{Var}[S_n]}{(\epsilon n)^2} \\
&= \frac{\sigma^2}{\epsilon^2 n}.
\end{aligned}
$$

Now as n goes to infinity, $\sigma^2/(\epsilon^2 n)$ goes to 0. Since we are bounding a probability, which is always at least 0, we must have

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0.$$

Chapter 4

Exercise 4.3:

(a) Determine the moment generating function for the binomial random variable $B(n, p)$.

(b) Let X be a $B(n, p)$ random variable and Y be a $B(m, p)$ random variable, where X and Y are independent. Use part (a) to determine the moment generating function of $X + Y$.

(c) What can we conclude from the form of the moment generating function of $X + Y$?

Solution to Exercise 4.3:

(a) We find using the binomial theorem

$$
\begin{aligned}
\mathbf{E}[e^{tX}] &= \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} e^{ti} \\
&= \sum_{i=0}^{n} \binom{n}{i} p e^{ti} (1-p)^{n-i} \\
&= (1 - p + p e^t)^n.
\end{aligned}
$$

(b) Since X and Y are independent,

$$
\begin{aligned}
\mathbf{E}[e^{t(X+Y)}] &= \mathbf{E}[e^{tX} e^{tY}] \\
&= \mathbf{E}[e^{tX}] \mathbf{E}[e^{tY}] \\
&= (1 - p + p e^t)^n (1 - p + p e^t)^m \\
&= (1 - p + p e^t)^{n+m}.
\end{aligned}
$$

(c) The moment generating function for $X + Y$ has the same form as the moment generating function for a $B(n + m, p)$ random variable. We conclude that the sum of a $B(n, p)$ random variable and $B(m, p)$ random variable gives a $B(n + m, p)$ random variable.

Exercise 4.6:

(a) In an election with two candidates using paper ballots, each vote is independently misrecorded with probability $p = 0.02$. Use a Chernoff bound to give an upper bound on the probability that more than 4 percent of the votes are misrecorded in an election of $1,000,000$ ballots.

(b) Assume that a misrecorded ballot always counts as a vote for the other candidate. Suppose that Candidate A received $510,000$ votes and Candidate B received $490,000$ votes. Use Chernoff bounds to upper bound the probability that Candidate B wins the election due to misrecorded ballots. Specifically, let X be the number of votes for Candidate A that are misrecorded and let Y be the number of votes for Candidate B that are misrecorded. Bound $\Pr((X > k) \cup (Y < \ell))$ for suitable choices of k and ℓ.

Solution to Exercise 4.6: For the first part, a variety of Chernoff bounds could be used. Applying Equation (4.1), for instance, gives that if Z is the number of misrecorded votes out of n total votes, then

$$
\Pr(Z \geq 2(0.02n)) \leq \left(\frac{e}{4}\right)^{0.02n}.
$$

With n being 1 million, the right hand side is $(e/4)^{20000}$, which is very small.

For the second part, as an example, in order for candidate B to win due to misrecording of ballots, at least one of the following must occur: either more than 15000 votes for A get misrecorded or less than 5000

votes for B get misrecorded. If both of these fail to occur, A wins, since A will obtain at least 510000 - 15000 + 5000 = 500000 votes. Hence bounding

$$\Pr((X > 15000) \cap \Pr(Y < 5000))$$

suffices. More generally, we could say that either more than $10000 + z$ votes for A get misrecorded or less than z votes for B get misrecorded, for any value of z; choosing the best value for z is then an optimization problem, and the best value could depend on the form of Chernoff bound used.

Exercise 4.7: Throughout the chapter we implicitly assumed the following extension of the Chernoff bound. Prove that it is true.

Let $X = \sum_{i=1}^{n} X_i$, where the X_i's are independent 0-1 random variables. Let $\mu = \mathbf{E}[X]$. Choose any μ_L and μ_H such that $\mu_L \leq \mu \leq \mu_H$. Then for any $\delta > 0$,

$$\Pr(X \geq (1+\delta)\mu_H) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{\mu_H}.$$

Similarly, for any $0 < \delta < 1$,

$$\Pr(X \leq (1-\delta)\mu_L) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu_L}.$$

Solution to Exercise 4.7:

(a) Let $t = \ln(1 + \delta) > 0$. By section 4.3 of the book, $\mathbf{E}[e^{tX}] \leq e^{(e^t-1)\mu} \leq e^{(e^t-1)\mu_H}$, where the second inequality follows from $e^t - 1 > 0$ for $t > 0$ and $\mu \leq \mu_H$. Following the proof of the normal Chernoff bound, we get:

$$
\begin{aligned}
\Pr(X \geq (1+\delta)\mu_H) &= \Pr(e^{tX} \geq e^{t(1+\delta)\mu_H}) \\
&\leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1+\delta)\mu_H}} \\
&\leq \frac{e^{(e^t-1)\mu_H}}{e^{t(1+\delta)\mu_H}} \\
&= \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{\mu_H}.
\end{aligned}
$$

(b) Let $t = \ln(1 - \delta) < 0$. By section 4.3 of the book, $\mathbf{E}[e^{tX}] \leq e^{(e^t-1)\mu} \leq e^{(e^t-1)\mu_L}$, where the second inequality follows from $e^t - 1 < 0$ for $t < 0$ and $\mu_L \leq \mu$. Following the proof of the normal Chernoff bound, we get:

$$
\begin{aligned}
\Pr(X \leq (1-\delta)\mu_L) &= \Pr(e^{tX} \geq e^{t(1-\delta)\mu_L}) \\
&\leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1-\delta)\mu_L}} \\
&\leq \frac{e^{(e^t-1)\mu_L}}{e^{t(1-\delta)\mu_L}} \\
&= \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right)^{\mu_L}.
\end{aligned}
$$

Exercise 4.9: Suppose that we can obtain independent samples $X_1, X_2, \ldots$, of a random variable X, and we want to use these samples to estimate $\mathbf{E}[X]$. Using t samples, we use $\sum_{i=1}^{t} X_i/t$ for our estimate of $\mathbf{E}[X]$. We want the estimate to be within $\epsilon \mathbf{E}[X]$ from the true value of $\mathbf{E}[X]$ with probability at least $1 - \delta$. We may not be able to use Chernoff's bound directly to bound how good our estimate is if X is not a $0 - 1$ random variable, and we do not know its moment generating function. We develop an alternative approach that requires only having a bound on the variance of X. Let $r = \frac{\sqrt{\mathbf{Var}[X]}}{\mathbf{E}[X]}$.

20

(a) Show using Chebyshev's inequality that $O(\frac{r^2}{\epsilon^2\delta})$ samples are sufficient to solve the above problem.

(b) Suppose that we only need a weak estimate that is within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$ with probability at least $3/4$. Argue that only $O(\frac{r^2}{\epsilon^2})$ samples are enough for this weak estimate.

(c) Show that by taking the median of $O(\log\frac{1}{\delta})$ weak estimates, we can obtain an estimate within $\epsilon\mathbf{E}[X]$ of $\mathbf{E}[X]$ with probability at least $1-\delta$. Conclude that we only need $O(\frac{r^2\log 1/\delta}{\epsilon^2})$ samples.

Solution to Exercise 4.9:

(a) We take $n = \lceil\frac{r^2}{\epsilon^2\delta}\rceil$ independent samples $X_1,\ldots,X_n$, and return $Z = \frac{1}{n}\sum_{i=1}^{n}X_i$ as our estimate.

By linearity of expectation, $\mathbf{E}[Z] = \mathbf{E}[X] = \mu$. Also, since the X_i are independent and all have the same distribution as X,

$$
\begin{aligned}
\mathbf{Var}[Z] &= \frac{1}{n^2}\mathbf{Var}\left[\sum_{i=1}^{n}X_i\right] \\
&= \frac{1}{n}\mathbf{Var}[X] \\
&= \frac{r^2\mu^2}{n}.
\end{aligned}
$$

Thus, by Chebyshev's Inequality,

$$
\begin{aligned}
\Pr(|Z-\mu| \geq \epsilon\mu) &\leq \frac{\mathbf{Var}[Z]}{\epsilon^2\mu^2} \\
&= \frac{r^2}{n\epsilon^2} \\
&\leq \delta.
\end{aligned}
$$

Hence $n = O(\frac{r^2}{\delta\epsilon^2})$ samples are sufficient.

(b) Just set $\delta = 1 - 3/4 = 1/4$ in part a. Hence we need $O(4\frac{r^2}{\epsilon^2}) = O(\frac{r^2}{\epsilon^2})$ samples.

(c) We run the algorithm from part b $m = \lceil 12\ln\frac{1}{\delta}\rceil$ times, obtaining independent estimates $S_1,\ldots,S_m$ for $\mathbf{E}[X]$. We then return the median of the S_i's, which we denote by M, as our new estimate for $\mathbf{E}[X]$.

Let Y_i be the 0-1 random variable that is 1 if and only if $|S_i-\mu| > \epsilon\mu$. By part b, $\mathbf{E}[Y_i] \leq 1/4$. Now, let $Y = \sum_{i=1}^{m}Y_i$. We see that $\mathbf{E}[Y] \leq m/4$ by linearity of expectations, and that Y is the sum of independent 0-1 random variables, so we can (and will) apply the Chernoff bound to Y.

Now, if $|M-\mu| > \epsilon\mu$, then by the definition of the median, $|S_i-\mu| > \epsilon\mu$ for at least $m/2$ of the S_i's, and so $Y \geq m/2 = (1+1)\frac{m}{4}$. Applying the Chernoff bound gives

$$
\begin{aligned}
\Pr(|M-\mu| > \epsilon\mu) &\leq \Pr(Y \geq (1+1)m/4) \\
&\leq e^{-\frac{1}{3}\cdot\frac{m}{4}\cdot 1^2} \\
&\leq e^{-\frac{1}{12}\cdot 12\ln\frac{1}{\delta}} \\
&\leq e^{\ln\delta} \\
&= \delta.
\end{aligned}
$$

Exercise 4.10: A casino is testing a new class of simple slot machines. Each game, the player puts in \$1, and the slot machine is supposed to return either \$3 to the player with probability 4/25, \$100 with proability 1/200, and nothing with all remaining probability. Each game is supposed to be independent of other games.

The casino has been surprised to find in testing that the machines have lost \$10,000 over the first millions game. Derive a Chernoff bound for the probability of this event. You may want to use a calculator or program to help you choose appropriate values as you derive your bound.

Solution to Exercise 4.10:

Let $X = \sum_{i=1}^{1000000} X_i$, where X_i is the winnings to the player from the ith game. With probability $1/200$, $X_i = 99$; With probability $4/25$, $X_i = 2$; and with probability $167/200$, $X_i = -1$. Note $\mathbf{E}[X_i] = -1/50$, so on average after 1 million games you would expect to lose 20000 dollars. It seems very odd for the casino to instead lose 10000 dollars.

A direct application of Markov's inequality gives, for $t > 0$,

$$
\begin{aligned}
\Pr(X \geq 10000) &= \Pr(e^{tX} \geq e^{10000t}) \\
&\leq \frac{\mathbf{E}[e^{tX}]}{e^{10000t}} \\
&= \frac{(\mathbf{E}[e^{tX_i}])^{1000000}}{e^{10000t}} \\
&= \frac{((167/200)e^{-t} + (4/25)e^{2t} + (1/200)e^{99t})^{1000000}}{e^{10000t}} \\
&= ((167/200)e^{-1.01t} + (4/25)e^{1.99t} + (1/200)e^{98.99t})^{1000000}
\end{aligned}
$$

Now we simply have to optimize to choose the right value of t. For $t = 0.0006$ the expression

$$
((167/200)e^{-1.01t} + (4/25)e^{1.99t} + (1/200)e^{98.99t})
$$

is approximately 0.9999912637. You can do a little better; for $t = 0.00058$ the expression is approximately 0.9999912508. The resulting bound on $\Pr(X \geq 10000)$ is approximately $1.6E - 4$.

Exercise 4.11: Consider a collection $X_1, \ldots, X_n$ of n independent integers chosen uniformly from the set $\{0, 1, 2\}$. Let $X = \sum_{i=1}^{n} X_i$ and $0 < \delta < 1$. Derive a Chernoff bound for $\Pr(X \geq (1 + \delta)n)$ and $\Pr(X \leq (1 - \delta)n)$.

Solution to Exercise 4.11:

We consider just the case of $\Pr(X \geq (1 + \delta)n)$. Using the standard Chernoff bound argument, for any $t > 0$,

$$
\begin{aligned}
\Pr(X \geq (1 + \delta)n) &= \Pr(e^{tX} \geq e^{t(1+\delta)n}) \\
&\leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1+\delta)n}} \\
&= \frac{(\mathbf{E}[e^{tX_i}])^n}{e^{t(1+\delta)n}} \\
&= \frac{((1 + e^t + e^{2t})/3)^n}{e^{t(1+\delta)n}} \\
&= \left(\frac{(1 + e^t + e^{2t})}{3e^{t(1+\delta)}} \right)^n.
\end{aligned}
$$

We therefore need to choose t minimize the expression

$$
\frac{(1 + e^t + e^{2t})}{3e^{t(1+\delta)}}.
$$

Sadly, there does not appear to be a simple form for t as a function of δ. For any specific δ, the best t can be found numerically. Using the choice $t = \ln(1 + \delta)$ used in other bounds, we have

$$
\Pr(X \geq (1 + \delta)n) \leq \left(\frac{1 + \delta + \delta^2/3}{(1 + \delta)^{(1+\delta)}} \right).
$$

Exercise 4.12: Recall that a function f is said to be *convex* if, for any x_1, x_2 and $0 \leq \lambda \leq 1$,

$$
f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).
$$

22

(a) Let Z be a random variable that takes on a (finite) set of values in the interval $[0, 1]$, and let $p = \mathbf{E}[Z]$. Define the Bernoulli random variable X by $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$. Show that $\mathbf{E}[f(Z)] \leq \mathbf{E}[f(X)]$ for any convex function f.

(b) Use the fact that $f(x) = e^{tx}$ is convex for any $t \geq 0$ to obtain a Chernoff-like bound for Z based on a Chernoff bound for X.

Solution to Exercise 4.12: Consider a collection $X_1, \ldots, X_n$ of n independent geometrically distributed random variables with mean 2. Let $X = \sum_{i=1}^{n} X_i$ and $\delta > 0$.

(a) Derive a bound on $\Pr(X \geq (1 + \delta)(2n))$ by applying the Chernoff bound to a sequence of $(1 + \delta)(2n)$ fair coin tosses.

(b) Directly derive a Chernoff bound on $\Pr(X \geq (1 + \delta)(2n))$ using the moment-generating function for geometric random variables.

(c) Which bound is better?

Solution to Exercise 4.12:

(a) Let Y denote the number of heads in $(1 + \delta)2n$ coin flips. We note that the events $\{X \geq (1+\delta)2n\}$ and $\{Y \leq n\}$ are equivalent, since the geometric random variables can be thought of as the number of flips until the next heads when flipping the coins. Now $\mathbf{E}[Y] = \frac{1}{2}(1 + \delta)2n = (1 + \delta)n = \mu$, so $n = \frac{1}{1+\delta}\mu = (1 - (1 - \frac{1}{1+\delta}))\mu = (1 - \frac{\delta}{1+\delta})\mu$. Using the Chernoff bound (4.5) we get

$$\Pr(Y \leq n) = \Pr(Y \leq (1 - \frac{\delta}{1+\delta})\mu) \leq= e^{-n\frac{\delta^2}{2(1+\delta)}}.$$

(b) From page 62 of the book, we have for a geometric random variable Z,

$$M_Z(t) = \frac{p}{1-p}((1 - (1-p)e^t)^{-1} - 1).$$

Plugging in $p = \frac{1}{2}$ we get

$$M_Z(t) = (1 - \frac{1}{2}e^t)^{-1} - 1 = \frac{\frac{1}{2}e^t}{1 - \frac{1}{2}e^t} = \frac{e^t}{2 - e^t}.$$

For $0 < t < \ln 2$, we get that

$$\Pr(X \geq (1 + \delta)2n) = \Pr(e^{tX} \geq e^{t(1+\delta)2n}) \leq \frac{\mathbf{E}[e^{tX}]}{e^{t(1+\delta)2n}}$$

$$\frac{M_X(t)}{e^{t(1+\delta)2n}} = \frac{\prod_{i=1}^{n} M_{X_i}(t)}{e^{t(1+\delta)2n}} = \frac{\left(\frac{e^t}{2-e^t}\right)^n}{e^{t(1+\delta)2n}} = e^{-nt(1+2\delta)}(2 - e^t)^{-nt}$$

Using calculus, we find the function $e^{-nt(1+2\delta)}(2 - e^t)^{-nt}$ is minimized for $t = \ln(1 + \frac{\delta}{\delta+1}) < \ln 2$. Finally we get that

$$\Pr(X \geq (1 + \delta)2n) \leq \min_{0 < t < \ln 2} e^{-nt(1+2\delta)}(2 - e^t)^{-nt} =$$

$$\left(1 + \frac{\delta}{\delta+1}\right)^{-n(1+2\delta)} \left(1 - \frac{\delta}{\delta+1}\right)^{-n}, \text{ setting } t = \ln(1 + \frac{\delta}{\delta+1}).$$

(c) Using calculus, we can show that the bound from the moment generating function is better under this derivation.

Exercise 4.13: Let $X_1, \ldots, X_n$ be independent Poisson trials such that $\Pr(X_i = 1) = p$. Let $X = \sum_{i=1}^{n} X_i$, so that $\mathbf{E}[X] = pn$. Let
$$F(x, p) = x \ln(x/p) + (1 - x) \ln((1 - x)/(1 - p)).$$

(a) Show that for $1 \geq x > p$,
$$\Pr(X \geq xn) \leq e^{-nF(x,p)}.$$

(b) Show that when $0 < x, p < 1$, $F(x, p) - 2(x - p)^2 \geq 0$. (Hint: take the second derivative of $F(x, p) - 2(x - p)^2$ with respect to x.)

(c) Using the above, argue that
$$\Pr(X \geq (p + \epsilon)n) \leq e^{-2n\epsilon^2}.$$

(d) Use symmetry to argue that
$$\Pr(X \geq (p - \epsilon)n) \leq e^{-2n\epsilon^2},$$

and conclude that
$$\Pr(|X - pn| \geq \epsilon n) \leq 2e^{-2n\epsilon^2}.$$

Solution to Exercise 4.13:

(a) In the proof of the basic Chernoff bound, we showed that
$$\begin{aligned}
M_{X_i}(t) &= \mathbf{E}[e^{tX_i}] \\
&= p_i e^t + (1 - p_i) \\
&= p e^t + (1 - p)
\end{aligned}$$

since in this case, $p_i = p$ always. It follows that $\mathbf{E}[e^{tX}] = M_X(t) = (pe^t + (1 - p))^n$. Now for any $t > 0$, by Markov's inequality,
$$\begin{aligned}
\Pr(X \geq xn) &= \Pr(e^{tX} \geq e^{txn}) \\
&\leq \mathbf{E}[e^{tX}]/e^{txn} \\
&= (pe^{t(1-x)} + (1 - p)e^{-tx})^n.
\end{aligned}$$

To find the appropriate value of t to minimize this quantity, we take the derivative of the expression inside the parentheses with respect to t. The derivative is:
$$p(1 - x)e^{t(1-x)} - (1 - p)xe^{-tx}.$$

It is easy to check that this is 0 (and gives a minimum) at
$$e^t = \frac{(1 - p)x}{p(1 - x)}, \text{ or } t = \ln \frac{(1 - p)x}{p(1 - x)},$$

and plugging in this value for t now gives the desired result, as
$$\begin{aligned}
\Pr(X \geq xn) &\leq \frac{(pe^t + (1 - p))^n}{e^{txn}} \\
&= \frac{((1 - p)x/(1 - x) + (1 - p))^n}{((1 - p)x/(p(1 - x)))^{xn}} \\
&= \frac{((1 - p)/(1 - x))^n}{((1 - p)x/(p(1 - x)))^{xn}} \\
&= \left(\frac{1 - p}{1 - x}\right)^{n(1-x)} \left(\frac{p}{x}\right)^{nx} \\
&= e^{-nF(x,p)}.
\end{aligned}$$

24

(b) The first derivative of $F(x,p) - 2(x-p)^2$ with respect to x is

$$\ln(x/p) - \ln((1-x)/(1-p)) - 4(x-p)$$

and the second derivative is (for $0 < x < 1$)

$$1/x + 1/(1-x) - 4 \geq 0.$$

As the first derivative is 0 when $x = p$ and the second derivative is non-negative we must have that $x = p$ gives a global minimum, so when $0 < x, p < 1$, $F(x,p) - 2(x-p)^2 \geq 0$.

(c) It now follows that

$$\Pr(X \geq (p+\epsilon)n) \leq \mathrm{e}^{-nF(p+\epsilon,p)} \leq \mathrm{e}^{-2n\epsilon^2}.$$

(d) Suppose we let $Y_i = 1 - X_i$ and $Y = \sum_{i=1}^{n} Y_i$. Since the Y_i are independent Poisson trials with $\Pr(Y_i = 1) = 1 - p$ we have

$$\begin{aligned}\Pr(X \leq (p-\epsilon)n) &= \Pr(Y \geq (1-p+\epsilon)n) \\ &\leq \mathrm{e}^{-2n\epsilon^2}\end{aligned}$$

by applying part (c) to Y. The result follows.

Exercise 4.19:

Solution to Exercise 4.19: It is worth noting that a function f is convex if it has a second derivative which is non-negative everywhere; this explains why e^{tx} is convex.

Let Z be a random variable that takes on a finite set of values $S \subset [0,1]$ with $p = \mathbf{E}[Z]$. Let X be a Bernoulli random variable with $p = \mathbf{E}[X]$. We show $\mathbf{E}[f(Z)] \leq \mathbf{E}[f(X)]$ for any convex function f. By definition, $f(s) \leq (1-s)f(0) + sf(1)$, so

$$\begin{aligned}\mathbf{E}[f(Z)] &= \sum_{s \in S} \Pr(Z = s)f(s) \\ &\leq \sum_{s \in S} \Pr(Z = s)(1-s)f(0) + \sum_{s \in S} \Pr(Z = s)f(1) \\ &= f(0)(1 - \mathbf{E}[Z]) + f(1)\mathbf{E}[Z] \\ &= f(0) \cdot (1-p) + f(1) \cdot p \\ &= \mathbf{E}[f(X)].\end{aligned}$$

For the second part, the first part implies that in the derivation of the Chernoff bound, we can upper bound $\mathbf{E}[\mathrm{e}^{tZ_i}]$ for any Z_i with distribution Z by $\mathbf{E}[\mathrm{e}^{tX_i}]$, where X_i is a Poisson trial with the same mean. Hence *any* Chernoff bound that holds for the sum of independent Poisson trials with mean p holds also for the corresponding sum of independent random variables all with distribution Z.

Exercise 4.21:

Consider the bit-fixing routing algorithm for routing a permutation on the n-cube. Suppose that n is even. Write each source node s as the concatenation of two binary strings a_s and b_s each of length $n/2$. Let the destination of s's packet be the concatenation of b_s and a_s. Show that this permutation causes the bit-fixing routing algorithm to take $\Omega(\sqrt{N})$ steps.

Solution to Exercise 4.21:

Assume that n is even and N is the total number of nodes (2^n). Take an arbitrary sequence of bits $b_1, \cdots, b_{\frac{n}{2}}$. Examine all the nodes on the hypercube of the form $(x_1, \cdots, x_{\frac{n}{2}-1}, \overline{b}_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$. Clearly we have 2^{n-1} choices for the x_i's so we have 2^{n-1} such nodes. Now by our choice of permutation, we know that the packet starting at $(x_1, \cdots, x_{\frac{n}{2}-1}, \overline{b}_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$ is routed to $(b_1, \cdots, b_{\frac{n}{2}}, x_1, \cdots, x_{\frac{n}{2}-1}, \overline{b}_{\frac{n}{2}})$. Since the bit fixing algorithm fixes one bit at a time from left to right, by the time the $\frac{n}{2} - 1$'th bit has been aligned the packet arrives at node $(b_1, \cdots, \overline{b}_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$ and hence to align the $\frac{n}{2}$'th bit the packet must take the edge to $(b_1, \cdots, b_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$. Hence the packets starting at the nodes $(x_1, \cdots, x_{\frac{n}{2}-1}, \overline{b}_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$ must all take the edge from $(b_1, \cdots, \overline{b}_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$ to $(b_1, \cdots, b_{\frac{n}{2}}, b_1, \cdots, b_{\frac{n}{2}})$. So $2^{n-1} = \sqrt{N}/2$ packets must use the same edge. Since only one packet can traverse an edge at a time, we need at least $\sqrt{N}/2 = \Omega(\sqrt{N})$ steps to complete the permutation.

Chapter 5

Exercise 5.1: For what values of n is $(1 + 1/n)^n$ within 1% of e? Within 0.0001% of e? Similarly, For what values of n is $(1 - 1/n)^n$ within 1% of 1/e? Within 0.0001%?

Solution to Exercise 5.1: We note that $(1+1/n)^n$ is increasing, which can be verified by calculus. Hence we need only find the first point where n reaches the desired thresholds. When $n = 50$, $(1+1/n)^n = 2.691588...$, and $e - (1+1/n)^n < 0.01e$. When $n = 499982$, $(1+1/n)^n = 2.718287911...$, and $e - (1+1/n)^n < 0.000001e$.

Similarly, $(1-1/n)^n$ is increasing. When $n = 51$, $(1+1/n)^n = 0.364243...$, and $1/e - (1+1/n)^n < 0.01/e$. When $n = 499991$, $(1 - 1/n)^n = 0.36787907...$, and $1/e - (1 + 1/n)^n < 0.000001/e$.

Exercise 5.4: In a lecture hall containing 100 people, you consider whether or not there are three people in the rooom who share the same birthday. Explain how to calculate this probability exactly, using the same assumptions as in our previous analysis.

Solution to Exercise 5.4:

Let us calculate the probability that there is no day with three birthdays. One approach is to set up a recurrence. Let $P(n, a, b)$ be the probability that when n people are each independently given a random birthday, there are a days where two people share a birthday and b days where just one person has a birthday. For generality let us denote the number of days in the year as m. Then we have $P(1, 0, 1) = 1.0$, and the recursion

$$P(n, a, b) = P(n - 1, a - 1, b + 1)\frac{b + 1}{m} + P(n - 1, a, b - 1)\frac{m - a - b + 1}{m}.$$

That is, if we consider the people one at a time, a new person either hits a day where someone else already has a birthday, or a new person is the first to have that birthday. This recurrence allows the calculation of the desired value.

Alternatively, we can express the probability as a summation. Let $k \geq 0$ be the number of days where two people share a birthday, and let $\ell \geq 0$ be the number of days where just one person has a birthday. Note that we must have $2k + \ell = 100$. There are $\binom{365}{k}$ ways to pick the k days, and then $\binom{365-k}{\ell}$ ways to pick the other ℓ days. Further, there are $100!/2^k$ ways to assign the 100 people to their birthdays; note we need to divide by 2^k to account for the fact that there are two equivalent ways of assigning two people to the same birthday for each of the k days. Each configuration has probability $(1/365)^{100}$ of occurring. Putting this all together, we have:

$$\sum_{k,\ell:k+2\ell=100;k,\ell\geq 0} \frac{365!100!}{k!\ell!(365 - k - \ell)!2^k 365^{100}}.$$

Exercise 5.7: Suppose that n balls are thrown independently and uniformly at random into n bins.

(a) Find the conditional probability that bin 1 has one ball given that exactly one ball fell into the first three bins.

(b) Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.

(c) Write an expression for the probability that bin 1 receives more balls than bin 2.

Solution to Exercise 5.7: For the first part, the ball is equally likely to fall in any of the three bins, so the answer must be $1/3$. For the second part, under the condition that bin 2 received no balls, the n balls are uniformly distributed over the remaining $n - 1$ bins; hence the conditional expectation must be $n/(n - 1)$. For the third part, it is easier to begin by writing an expression for the probability the two bins receive the same number of balls. This is:

$$P = \sum_{k=0}^{n/2} \binom{n}{k; k; n - k} \left(\frac{1}{n}\right)^{2k} \left(1 - \frac{2}{n}\right)^{n-2k}.$$

By symmetry, the probability we seek is $(1 - P)/2$.

Exercise 5.9: Consider the probability that every bin receives exactly one ball when n balls are thrown randomly into n bins.

(a) Give an upper bound on this probability using the Poisson approximation.

(b) Determine the *exact* probability of this event.

(c) Show that these two probabilities differ by a multiplicative factor that equals the probability that a Poisson random variable with parameter n takes on the value n. Explain why this is implied by Theorem 5.6.

Solution to Exercise 5.9:

(a) With the Poisson approximation, we have each bin independently obtains a number of balls that is Poisson distributed with mean 1. Hence the probability each bin obtains one ball is $(e^{-1})^n = e^{-n}$. Using Theorem 5.7, for example, would give an upper bound of $e\sqrt{n}e^{-n}$; more tightly, delving into Theorem 5.7, we would have an upper bound of $n!/n^n$.

(b) This probability is just $n!/n^n$. There are $n!$ ways of distributing the balls one per bin, and n^n total ways of distributing the balls.

(c) The first two parts show that the probability from the Poisson approximation, namely e^{-n}, and the exact probability, $n!/n^n$, differ by a factor corresponding exactly to the probability that a Poisson random variable with parameter n takes on the value n. Looking at Theorem 5.6, we see that this has to be the case. Note that for the Poisson approximation, we can first require that n total balls be thrown; otherwise there is no chance of having exactly one ball in each bin. Conditioned on this event, Theorem 5.6 tells us that the distribution will be exactly that of throwing balls into bins. Hence the difference between the Poisson approximation and throwing balls into bins in this case is exactly the initial conditioning on the number of balls being n with the Poisson approximation, giving the precise factor between the probabilities.

Exercise 5.11: The following problem models a simple distributed system wherein agents contend for resources but "back off" in the face of contention. Balls represent agents, and bins represent resources.

The system evolves over rounds. Every round, balls are thrown independently and uniformly at random into n bins. Any ball that lands in a bin by itself is *served* and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we finish when every ball is served.

(a) If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?

(b) Suppose that every round the number of balls served was exactly the expected number of balls to be served. Show that all the balls would be served in $O(\log \log n)$ rounds. (Hint: if x_j is the expected number of balls left after j rounds, show and use that $x_{j+1} \le x_j^2/n$.)

Solution to Exercise 5.11:

(a) If there are b balls at the start of a round, the probability a ball lands alone in a bin is

$$ p = \left(1 - \frac{1}{n}\right)^{b-1}. $$

Hence the expected number of balls in the next round is $n(1 - p)$. Note that this is approximately $n(1 - e^{-b/n})$.

(b) Consider any ball in round $j + 1$. Considering all the other balls, at most x_j of the bins have at least one ball in it, since there are at most x_j balls being placed in the round. Hence the probability any ball is placed into a bin where there is some other ball is at most x_j/n. It follows that the expected number of balls that are left after $j + 1$ rounds is at most $x_j \cdot (x_j/n) = x_j^2/n$.

27

From the calculations in part (a), it is easy to check that after two rounds, the expected number of balls remaining is at most $n/2$. It follows inductively that for $j \geq 2$

$$x_j \leq 2^{-2^{j-2}} n.$$

After $j = \log \log n + 3$ rounds, $x_j < 1$; since x_j is supposed to be integer-valued, $O(\log \log n)$ rounds suffice. (Alternatively, after $O(\log \log n)$ rounds of following the expectation, we are down to 1 ball, at which point only 1 more round is necessary.)

Exercise 5.16: Let G be a random graph generated using the $G_{n,p}$ model.

(a) A *clique* of k vertcies in a graph is a subset of k vertices such that all $\binom{k}{2}$ edges between these vertices lie in the graph. For what value of p, as a function of n, is the expected number of cliques of five vertices in G equal to 1?

(b) A $K_{3,3}$ graph is a complete bipartite graph with three vertices on each side. In other words, it is a graph with six vertices and nine edges; the six distinct vertices are arranged in two groups of three, and the nine edges connect each of the nine pairs of vertices with one vertex in each group. For what value of p, as function of n, is the expected number of $K_{3,3}$ subgraphs in G equal to one?

(c) For what value of p, as a function of n, is the expected number of Hamiltonian cycles in the graph equal to 1?

Solution to Exercise 5.16:

(a) There are $\binom{n}{5}$ possible cliques of size 5. Each possible clique has $\binom{5}{2} = 10$ edges. Hence the expected number of cliques is $\binom{n}{5}p^{10}$, and this expectation equals 1 when $p = \sqrt[10]{\binom{n}{5}^{-1}}$.

(b) There are $\frac{1}{2}\binom{n}{3} \cdot \binom{n-3}{3}$ possible ways of choosing the six vertices of the $K_{3,3}$; first we choose three vertices for one side, then we choose three of the remaining vertices for the other side. (We have to divide by two because each $K_{3,3}$ is counted twice, as each set of vertices could show up on either side.) As there are 9 edges in a $K_{3,3}$, the expected number that appear is $\frac{1}{2}\binom{n}{3} \cdot \binom{n-3}{3}p^9$. For this to equal 1, we need $p = \sqrt[9]{2\binom{n}{3}^{-1}\binom{n-3}{3}^{-1}}$.

(c) There are $(n-1)!/2$ different Hamiltonian cycles in an (undirected) graph. Hence the expected number of Hamiltonian cycles is $p^n(n-1)!/2$, and this expectation is 1 when $p = \sqrt[n]{2/(n-1)!}$.

Exercise 5.21: In hashing with *open addressing*, the hash table is implemented as an array, and there are no linked lists or chaining. Each entry in the array either contains one hashed item or it is empty. The hash function defines for each key k a *probe sequence* $h(k,0), h(k,1), \ldots$ of table locations. To insert the key k, we examine the sequence of table locations in the order defined by the key's probe sequence until we find an empty location and then insert the item at that position. When searching for an item in the hash table, we examine the sequence of table locations in the order defined by the key's probe sequence until either the item is found, or we have found an empty location in the sequence. If an empty location is found, it means that the item is not present in the table.

An open address hash table with $2n$ entries is used to store n items. Assume that the table location $h(k,j)$ is uniform over the $2n$ possible table locations and that all $h(k,j)$ are independent.

(a) Show that under these conditions the probability an insertion requires more than k probes is at most 2^{-k}.

(b) Show that for $i = 1, 2, \ldots, n$ the probability that the i-th insertion requires more than $2\log n$ probes is at most $1/n^2$.

Let the random variable X_i denote the number of probes required by the i-th insertion. You have shown in the previous question that $\Pr(X_i > 2\log n) \leq 1/n^2$. Let the random variable $X = \max_{1 \leq i \leq n} X_i$ denote the maximum number of probes required by any of the n insertions.

(c) Show that $\Pr(X > 2\log n) \le 1/n$.

(d) Show that the expected length of the longest probe sequence is $E[X] = O(\log n)$.

Solution to Exercise 5.21:

(a) Since the hash table contains $2n$ entries and at any point at most n elements, the probability that a given probe fails to find an empty entry is at most $1/2$. Therefore the probability that an insertion needs k probes is at most 2^{-k}.

(b) We apply part (a) with $k = 2\log n$.

(c) We apply the union bound to part (b).

(d) We find

$$\mathbf{E}[X] = \Pr(X \le 2\log n) \cdot \mathbf{E}[X \mid X \le 2\log n] + \Pr(X > 2\log n) \cdot \mathbf{E}[X \mid X > 2\log n]$$

$$\le 2\log n + \frac{1}{n}\left(2\log n + \sum_{k=2\log n}^{\infty} \Pr(X \ge k \mid X > 2\log n)\right)$$

$$\le 2\log n + \frac{1}{n}\left(2\log n + \sum_{k=1}^{\infty} \Pr(X \ge k)\right)$$

$$\le 2\log n + \frac{1}{n}\left(2\log n + n\sum_{k=1}^{\infty} 2^{-k}\right)$$

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

$$\le 2\log n + 2,$$

for sufficiently large n.

Exercise 5.22: Bloom filters can be used to estimate set differences. Suppose you have a set X and I have a set Y, both with n elements. For example, the sets might represent our 100 favorite songs. We both create Bloom filters of our sets, using the same number of bits m and the same k hash functions. Determine the expected number of bits where our Bloom filters differ as a function of m, n, k, and $|X \cap Y|$. Explain how this could be used as a tool to find people with the same taste in music more easily than comparing lists of songs directly.

Solution to Exercise 5.22:

Consider the ith position in the filter. For it to be set to 1 by the set X but not by the set Y, it must be set by the hash of some element in $X - (X \cap Y)$, but not set by the hash of any element in Y. This probability is

$$\left(1 - \frac{1}{m}\right)^{k|Y|}\left(1 - \left(1 - \frac{1}{m}\right)^{k(|X|-|X\cap Y|)}\right).$$

Note that $|X| = |Y| = n$, so we can write this as

$$\left(1 - \frac{1}{m}\right)^{kn}\left(1 - \left(1 - \frac{1}{m}\right)^{k(n-|X\cap Y|)}\right).$$

The probability that the bit is set to 1 by the Y but not by the set X is the same. Hence the expected number of bits set differently is just

$$2m\left(1 - \frac{1}{m}\right)^{kn}\left(1 - \left(1 - \frac{1}{m}\right)^{k(n-|X\cap Y|)}\right).$$

This approach could be used to estimate how much two people share the same taste in music; the higher the overlap between the Bloom filters of their 100 favorite songs, the larger the intersection in their favorite songs list. Since the Bloom filters are like compressed versions of the lists, comparing them may be much faster than comparing song lists directly.

Exercise 5.23:

Suppose that we wanted to extend Bloom filters to allow deletions as well as insertions of items into the underlying set. We could modify the Bloom filter to be an array of counters instead of an array of bits. Each time an item is inserted into a Bloom filter, the counters given by the hashes of the item are increased by one. To delete an item, one can simply decrement the counters. To keep space small, the counters should be a fixed length, such as 4 bits.

Explain how errors can arise when using fixed-length counters. Assuming a setting where one has at most n elements in the set at any time, m counters, k hash functions, and counters with b bits, explain how to bound the probability that an error occurs over the course of t insertions or deletions.

Solution to Exercise 5.23: This variation of the Bloom filter is generally referred to as a counting Bloom filter in the literature.

Problems can occur with counting Bloom filters if there is counter overflow. An overflow occurs whenever 2^b of the nk hashes fall in the same counter. We can use the Poisson approximation to bound the probability of counter overflow, treating each hash as a ball and each counter as a bin. (By Theorem 5.10, the probabilities we compute via the Poisson approximation are within a factor of 2 of the exact probabilities, since counter overflow increases with the number of hashes.) The Poisson approximation says that the probability single counter does not overflow at any point in time is at least

$$\sum_{j=0}^{2^b-1} \frac{e^{-kn/m}(kn/m)^j}{j!}.$$

(The probability could be even greater if there were less than n elements at that time.) The probability that no counter overflows is then at least

$$\left(\sum_{j=0}^{2^b-1} \frac{e^{-kn/m}(kn/m)^j}{j!}\right)^m.$$

Including the factor of 2 from the approximation, the probability of any counter overflow is bounded by

$$2\left(1 - \left(\sum_{j=0}^{2^b-1} \frac{e^{-kn/m}(kn/m)^j}{j!}\right)^m\right).$$

Finally, the above is for any fixed moment in time, so by a union bound, we can bound the probability over all time steps by

$$2t\left(1 - \left(\sum_{j=0}^{2^b-1} \frac{e^{-kn/m}(kn/m)^j}{j!}\right)^m\right).$$

As an example, when $b = 4$, $n = 10000$, $m = 80000$, and $k = 5$, we have a bound of approximately $0.23 \cdot 10^{-11} \cdot t$, so one can go many steps without an overflow.

Solution to the Exploratory Assignment

For better or worse, a great deal about the assignment can be found in the paper "An Experimental Assignment on Random Processes", available at Michael Mitzenmacher's home page. This makes useful reading for an instructor in preparation for the assignment; however, an instructor should be on the lookout for students copying from this decscription.

A key insight into the assignment is that, for each pair of sibling leaves at the bottom of the tree, at least one of the two siblings must arrive in order for all the node to be marked.

The first process behaves like a coupon collector's problem: each pair of sibling leaves must have one of its leaves marked. The probability any given pair is hit on any specific step is $2/N$, and there are $Q = (N+1)/2$

pairs of leaves. Let X_i be the number of nodes sent after the $(i-1)$st pair is hit until the ith pair is hit, and let $X = \sum X_i$. Then

$$\mathbf{E}[X] = \frac{N}{2Q} + \frac{N}{2(Q-1)} + \ldots + \frac{N}{2} = \frac{N}{2} \sum_{i=1}^{Q} \frac{1}{i} = \Omega(N \log Q) = \Omega(N \log N),$$

and this is a lower bound on the expected total nodes sent.

For the second process, the behavior is similar to the birthday paradox. Let us first consider just the $Q = (N+1)/2$ leaf nodes, in the order they arrive. What is the probability some pair of sibling leaf nodes both lie in the last $L = c\sqrt{Q}$ leaf nodes (for some constant c)? Reversing the order, so that we look at these Q nodes from back to front, this is equivalent to asking whether we have some pair in the first $c\sqrt{Q}$ nodes. The probability that no such pair occurs is just

$$\prod_{i=1}^{L} \left(1 - \frac{i-1}{Q-i+1}\right) \le e^{\sum_{i=1}^{L}(i-1)/(Q-i)} \le e^{-\sum_{i=1}^{L}(i-1)/(Q-i)} \le e^{-L^2/(Q-L)} \le e^{-c^2}.$$

Hence the probability that last $L = c\sqrt{Q}$ nodes contain a pair of sibling leaves it at least some constant.

Now let $c = 1$, and note that within the last $2\sqrt{N}$ sent nodes, with probability at least $1/2$, at least $1/2$ of these nodes are leaf nodes (just by symmetry). The result therefore follows immediately, using say $c = 1$.

The behavior of the third process is completely deterministic. It takes exactly $(N+1)/2$ nodes to arrive in all cases. This is easily proved by induction on n, taking cases on whether one of the $(N+1)/2$ nodes that arrives is the root or not. Alternatively, there is a nice interpretation. Let us think of each leaf node as a variable, and each internal node as the sum of variables in its subtree. Interpret being sent a node as being given a value for the corresponding variable or sum of variables. We have $(N+1)/2$ variables; once we have $(N+1)/2$ values, we have $(N+1)/2$ linearly independent equations in these $(N+1)/2$ variables, which can then be solved, giving all the values and allowing us to fill the tree. The marking actions on the tree correspond exactly to solving for all possible variables (or sums of variables) given the values present.

Chapter 6

Exercise 6.2:

(a) Prove that, for every integer n, there exists a coloring of the edges of the complete graph K_n by two colors so that the total number of monochromatic copies of K_4 is at most $\binom{n}{4}2^{-5}$.

(b) Give a randomized algorithm for finding a coloring with at most $\binom{n}{4}2^{-5}$ monochromatic copies of K_4 that runs in expected time polynomial in n.

(c) Show how to construct such a coloring deterministically in polynomial time using the method of conditional expectations.

Solution to Exercise 6.2:

(a) Suppose we just color the edges randomly, independently for each edge. Enumerate the $\binom{n}{4}$ copies of K_4 in the graph in any order. Let $X_i = 1$ if the ith copy is monochromatic, and 0 otherwise. Then $X = \sum X_i$, and by linearity $\mathbf{E}[X] = \sum \mathbf{E}[X_i]$. Each copy of K_4 has 6 edges, and is therefore monochromatic with probability 2^{-5}, so we have $\mathbf{E}[X] = \binom{n}{4}2^{-5}$. It follows that there is at least one coloring with at least this many monochromatic K_4.

(b) A randomized algorithm would simply color each edge randomly and count the number of monochromatic copies of K_4. Such counting takes time at most $O(n^4)$, as there are $\binom{n}{4}$ copies to check. For convenience, let us assume n is large enough so that $\binom{n}{4}2^{-5}$ is an integer. The probability p that each randomized trial succeeds then satisifies:

$$p\binom{n}{4} + (1-p)\left(\binom{n}{4}2^{-5} - 1\right) \geq \binom{n}{4}2^{-5},$$

or

$$p \geq \frac{1}{\frac{31/32}{\binom{n}{4}+1}}.$$

Hence it takes on average at most $O(n^4)$ trials to have one with the required monochromatic coloring. We conclude the randomized guess-and-check algorithm takes time at most $O(n^8)$.

(c) Consider the edges in some arbitrary order. We color the edges one at a time. Each time we color an edge, we consider both possible colors, and for each coloring we compute the expected number of monochromatic K_4 that will be obtained if we color the remaining edges randomly. We give the edge the color with the greater expectation. This can be computed in at most $O(n^4)$ time each time we color an edge, simply by checking each copy of K_4 and determining the probability it will still be colored monochromatically. In fact, this computation can be done much more quickly; for example, we need really only consider and compare the expectation for copies of K_4 that include the edge to be colored, which reduces the computation to $O(n^2)$ for each edge colored. As there are m edges, this gives an $O(n^4m)$ (or $O(n^2m)$) algorithm. At each step, the expected number of monochromatic K_4 is at least $\binom{n}{4}2^{-5}$, so the algorithm terminates with the desired coloring.

Exercise 6.3: Given an n-vertex undirected graph $G = (V, E)$, consider the following method of generating an independent set. Given a permutation σ of the vertices, define a subset $S(\sigma)$ of the vertices as follows: for each vertex i, $i \in S(\sigma)$ if and only if no neighbor j of i precedes i in the permutation σ.

(a) Show that each $S(\sigma)$ is an independent set in G.

(b) Suggest a natural randomized algorithm to produce σ for which you can show that the expected cardinality of $S(\sigma)$ is

$$\sum_{i=1}^{n} \frac{1}{d_i + 1},$$

where d_i denotes the degree of vertex i.

(c) Prove that G has an independent set of size at least $\sum_{i=1}^{n} \frac{1}{d_i+1}$.

Solution to Exercise 6.3:

(a) Supppose $i, j \in S(\sigma)$ where i and j are neighbors. But if i precedes j in σ then j cannot be in $S(\sigma)$, and similarly if j precedes i. In either case we have a contradiction, so no two elements of $S(\sigma)$ can be neighbors, and $S(\sigma)$ forms an independent set.

(b) We choose σ uniformly at random. We now define X_i by $X_i = 1$ if $i \in S(\sigma)$ and $X_i = 0$ otherwise. We have $X_i = 1$ if and only if i precedes all of its neighbors in σ. Now consider the subset of elements S_i consisting of i and all of its neighbors. Since the permutation is chosen uniformly at random, each element of S_i is equally likely to be the first element of S_i to appear in σ. Since $|S_i| = d_i + 1$, we see that $\Pr(X_i = 1) = \frac{1}{d_i+1}$. As we can represent $|S(\sigma)| = \sum_{i=1^n} X_i$, we have $E[|S(\sigma)|] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{d_i+1}$.

(c) We have shown that for a random permutation σ, $E[|S(\sigma)|] = \sum_{i=1}^{n} \frac{1}{d_i+1}$. We can therefore conclude that there exists a permutation π such that $S(\pi) \geq \sum_{i=1}^{n} \frac{1}{d_i+1}$, and therefore there is a independent set $S(\pi)$ of the desired size.

Exercise 6.4: Consider the following 2-player game. The game begins with k tokens, placed at the number 0 on the integer number line spanning $[0, n]$. Each round, one player, called the *chooser*, selects two disjoint and nonempty sets of tokens A and B. (The sets A and B need not cover all the remaining tokens; they only need to be disjoint.) The second player, called the *remover*, takes all the tokens from one of the sets off the board. The tokens from the other set all move up one space on the number line from their current position. The chooser wins if any token ever reaches n. The remover wins if the chooser finishes with one token that has not reached n.

(a) Give a winning strategy for the chooser when $k \geq 2^n$.

(b) Use the probabilistic method to show that there must exist a winning strategy for the remover when $k < 2^n$.

(c) Explain how to use the method of conditional expectations to derandomize the winning strategy for the remover when $k < 2^n$.

Solution to Exercise 6.4:

(a) At each round, the chooser splits the tokens up into two equal size sets (or sets as equal as possible) A and B. By induction, after j rounds, the chooser has at least 2^{n-j} tokens at position j, and hence has at least one token at position n after n rounds.

(b) Suppose that the remover just chooses a set to remove randomly (via an independent, fair coin flip) each time. Let X be the number of tokens that ever reach position n when the remover uses this strategy. Let $X_m = 1$ if the mth token ever reaches position n and 0 otherwise, so $X = \sum_{i=1}^{k} X_m$. For the mth token to reach position n, it has to be moved forward n times; each time the chooser puts it in a set, it is removed with probability $1/2$. Hence the probability it reaches position n is at most $1/2^n$, from which we have $\mathbf{E}[X_m] \leq 1/2^n$ and thus $\mathbf{E}[X] \leq k/2^n < 1$. Since a random strategy yields on average less than 1 token that reaches position n, there must be a strategy that yields 0 tokens that reach position n.

(c) At any point in the game, let k_i be the number of tokens at position i. With the randomized strategy of part (b), the expected number of tokens that reach position n is $\sum_{i=0}^{n} k_i 2^{i-n}$. At each round, the remover should remove the set that minimizes this expectation after the set is removed. Inductively, after each round, the expected number of tokens that will reach position n is less than 1. Hence, regardless of the strategy of the chooser, the remover guarantees that 0 tokens reach position n.

Exercise 6.5: We have shown using the probabilistic method that, if a graph G has n nodes and m edges, there exists a partition of the n nodes into sets A and B so that at least $m/2$ edges cross the partition. Improve this result slightly: show that there exists a partition so that at least $\frac{mn}{2n-1}$ edges cross the partition.

Solution to Exercise 6.5: Instead of partitioning the vertices by placing each randomly, uniformly choose a partition with $n/2$ edges on each side if n is even, or $(n-1)/2$ on one side and $(n+1)/2$ on the other if n is odd. In the first case, for any edge (x,y) there are $2\binom{n-2}{(n-2)/2}$ partitions of the vertices where x and y are on opposite sides (2 ways to position x and y, $\binom{n-2}{(n-2)/2}$ ways to partition the other vertices). Hence the probability the edge crosses the cut is $2\binom{n-2}{(n-2)/2}/\binom{n}{n/2} = n/(2n-2)$. By linearity of expectations the expected number of edges crossing the cut is $\frac{mn}{2n-2}$. The second case is similar, but the the probability the edge crosses the cut is $2\binom{n-2}{(n-1)/2}/\binom{n}{(n+1)/2} = (n+1)/(2n)$, and the expected number of edges crossing the cut is $\frac{m(n+1)}{2n}$. In both cases the result is better than $\frac{mn}{2n-1}$.

Exercise 6.9: A *tournament* is a graph on n vertices with exactly one directed edge between each pair of vertices. If vertices represent players, each edge can be thought of as the result of a match between the two players; the edge points to the winner. A *ranking* is an ordering on the n players from best to worst (ties are not allowed). Given the outcome of a tournament, one might wish to determine a ranking of the players. A ranking is said to *disagree* with a directed edge from y to x if y is ahead of x in the ranking (since x beat y in the tournament).

  (a) Prove that, for every tournament, there exists a ranking that disagrees with at most 50% of the edges.

  (b) Prove that, for sufficiently large n, there exists a tournament such that every ranking disagrees with at least 49% of the edges in the tournament.

Solution to Exercise 6.9:

  (a) For a ranking chosen uniformly at random, each edge individually agrees with ranking with probability $1/2$. Hence the expected number of edges that agree is $1/2$ the number of edges, and thus there must exist a ranking that agrees on at least $1/2$ the edges.

  (b) Choose a random tournament, where the direction of each edge is chosen independently, and consider any fixed ranking. Let X be the number of edges that agree with ranking. Clearly $\mathbf{E}[X] = \binom{n}{2}/2$. Since each edge is chosen independently, we must have by Chernoff bounds (such as, for example, Corollary 4.9 on page 70)
  $$\Pr\left(X \geq 0.51\binom{n}{2}\right) \leq e^{-\binom{n}{2}/10000}.$$

  Now let us take a union bound over *all* $n!$ different rankings. The probability that any of these rankings agree with at least $0.51\binom{n}{2}$ is at most
  $$n!e^{-\binom{n}{2}/10000}.$$

  Since $n! \leq n^n = e^{n \ln n}$, we see that for large enough n, this expression goes to 0. Hence, for large enough n, a random tournament is very likely to have no ranking that agrees in 51 percent of the matches, which gives the result.

Exercise 6.10: A family of subsets $\mathcal{F}$ of $\{1, 2, \ldots, n\}$ is called an *antichain* if there is no pair of sets A and B in $\mathcal{F}$ satisfying $A \subset B$.

  (a) Give an example of $\mathcal{F}$ where $|\mathcal{F}| = \binom{n}{\lfloor n/2 \rfloor}$.

  (b) Let f_k be the number of sets in $\mathcal{F}$ with size k. Show that
  $$\sum_{k=0}^{n} \frac{f_k}{\binom{n}{k}} \leq 1.$$

  (*Hint*: Choose a random permutation of the numbers from 1 to n, and let $X_k = 1$ if the first k numbers in your permutation yield a set in $\mathcal{F}$. If $X = \sum_{k=0}^{n} X_k$, what can you say about X?)

(c) Argue that $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$ for any antichain $\mathcal{F}$.

Solution to Exercise 6.10:

(a) The example is just to take all subsets of $\lfloor n/2 \rfloor$ elements.

(b) For $X = \sum_{k=0}^{n} X_k$, we must have $X \leq 1$, since only one fo the X_i can be 1 by definition. (If two of the X_i equal 1, we have two sets, one a subset of the other, both in the antichain. This contradicts the definition.) Hence $\mathbf{E}[X] \leq 1$. But by linearity

$$1 \geq \mathbf{E}[X] = \sum_{k=0}^{n} \mathbf{E}[X_k] = \sum_{k=0}^{n} \Pr(X_k = 1) = \sum_{k=0}^{n} \frac{f_k}{\binom{n}{k}},$$

giving the result.

(c) As

$$1 \geq \sum_{k=0}^{n} \frac{f_k}{\binom{n}{k}} \geq \sum_{k=0}^{n} \frac{f_k}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{\mathcal{F}}{\binom{n}{\lfloor n/2 \rfloor}},$$

the result follows.

Exercise 6.13: Consider a graph in $G_{n,p}$, with $p = \frac{c \ln n}{n}$. Use the second moment method or the conditional expectation inequality to prove that if $c < 1$ then, for any constant $\epsilon > 0$ and for n sufficiently large, the graph has isolated vertices with probability at least $1 - \epsilon$.

Solution to Exercise 6.13: First, let us determine the expected number of isolated vertices. The probability a vertex is isolated is the probability that none of its adjacent edges appear, which is

$$\left(1 - \frac{c \ln n}{n}\right)^n \approx e^{-c \ln n} = n^{-c}.$$

Note that here and throughout we use the appoximation $1 - x \approx e^{-x}$ when x is small; for n sufficiently large, this approximation is suitable for this argument, as the error affects only lower order terms, so we may think of the approximation as an equality up to these lower order terms. The expected number of isolated vertices is therefore approximately n^{1-c}, which is much greater than 1 for $c < 1$.

Now consider using the conditional expectation inequality. Let X_i be 1 if the ith vertex is isolated and 0 otherwise, and let $X = \sum X_i$. We have found that $\Pr(X_i = 1) \approx n^{-c}$. Also,

$$\mathbf{E}[X \mid X_i = 1] = 1 + \sum j \neq i \Pr(X_j = 1 \mid X_i = 1).$$

Now, the effect that vertex i is isolated does not affect X_j substantially; the probability that the jth vertex is isolated (for $j \neq i$) after this conditioning is

$$\Pr(X_j = 1 \mid X_i = 1) = \left(1 - \frac{c \ln n}{n}\right)^{n-1} \approx n^{-c}.$$

Again, here we use approximation to signify the difference is only in lower order terms. From the conditional expectation inequality, and using our approximations, we have

$$\Pr(X > 0) \geq \frac{n^{1-c}}{1 + (n - 1)n^{-c}} \geq 1 - \epsilon$$

for any constant ϵ when n is sufficiently large.

Exercise 6.14: Consider a graph in $G_{n,p}$, with $p = \frac{1}{n}$. Let X be the number of triangles in the graph, where a triangle is a clique with three edges. Show that

$$\Pr(X \geq 1) \leq 1/6$$

and that
$$\lim_{n \to \infty} \Pr(X \geq 1) \geq 1/7.$$

(*Hint:* use the conditional expectation inequality.)

Solution to Exercise 6.14:

The expectation $\mathbf{E}[X]$ is given by

$$\mathbf{E}[X] = \binom{n}{3} \left(\frac{1}{n}\right)^3 \leq \frac{1}{6}.$$

Since X takes on values on the non-negative integers, it follows that $\Pr(X > 0) \leq 1/6$.

To use the conditional expectation inequality, we enumerate the $\binom{n}{3}$ possible triangles in some arbitrary way. We let $X_i = 1$ if the ith triangle appears in the graph, and $X_i = 0$ otherwise. Clearly $X = \sum_i X_i$, and $\Pr(X_i = 1) = 1/n^3$. Now

$$\mathbf{E}[X \mid X_i = 1] = 1 + \sum_{j \neq i} \Pr(X_j = 1 \mid X_i = 1).$$

There are $\binom{n-3}{3}$ ways of choosing a triangle so that it is disjoint from X_i. In this case the edges of the triangles are independent and $\Pr(X_j = 1 \mid X_i = 1) = 1/n^3$. Similarly, there are $3\binom{n-3}{2}$ triangles that share a vertex but no edges with X_i, and for these triangles $\Pr(X_j = 1 \mid X_i = 1) = 1/n^3$. Finally, there are $3\binom{n-3}{1}$ triangles that share two vertices and hence one edge with X_i, and for these triangles $\Pr(X_j = 1 \mid X_i = 1) = 1/n^2$. Putting this together in the conditional expectation inequality we have

$$\Pr(X > 0) \geq \binom{n}{3} \frac{\frac{1}{n^3}}{1 + \binom{n-3}{3}\frac{1}{n^3} + 3\binom{n-3}{2}\frac{1}{n^3} + 3\binom{n-3}{1}\frac{1}{n^2}}.$$

Taking the limit now gives the desired result.

Exercise 6.16: Use the Lovasz local lemma to show that, if

$$4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}} < 1,$$

then it is possible to color the edges of K_n so that it has no monochromatic K_k subgraph.

Solution to Exercise 6.16: To use the Lovasz local lemma, we need to set up a group of events so that $4dp \leq 1$, where d is the degree of the dependency graph, and p is an upper bound on the probability of each event. There will be $\binom{n}{k}$ events, one event for each K_k subgraph, with the event being that the subgraph is monochromatic. Then $p = 2^{1-\binom{k}{2}}$. Two events are dependent if they share an edge. Each edge is a part of $\binom{n-2}{k-2} < \binom{n}{k-2}$ subgraphs of type K_k, so each K_k subgraph shares an edge with fewer than $\binom{k}{2}\binom{n}{k-2}$ other K_k subgraphs. This gives our bound on d, and by the Lovasz local lemma, we have a coloring where no K_k graph is monochromatic.

Chapter 7

Exercise 7.2: Consider the two-state Markov chain with the following transition matrix.

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Find a simple expression for $P_{0,0}^t$.

Solution to Exercise 7.2:

$$P_{0,0}^t = \frac{1 + (2p-1)^t}{2}.$$

This is easily shown by induction. See also the related Exercise 1.11, which considers the equivalent problem in a different way.

Exercise 7.12: Let X_n be the sum of n independent rolls of a fair die. Show that for any $k \geq 2$,

$$\lim_{n \to \infty} \Pr(X_n \text{ is divisible by } k) = \frac{1}{k}.$$

Solution to Exercise 7.12: Let $Z_1, Z_2, \ldots$ be the sequence defined as follows: $Z_j = (X_j \bmod k)$. Notice that the Z_j give a Markov chain! It is easy to check that the stationary distributon of this chain is uniform over $0, 1, \ldots, k-1$, using for example Definition 7.8. Therefore we have by Theorem 7.7 that

$$\lim_{n \to \infty} \Pr(Z_n = 0) = 1/k,$$

which is equivalent to the desired result.

Exercise 7.17: Consider the following Markov chain, which is similar to the 1-dimensional random walk with a completely reflecting boundary at 0. Whenever position 0 is reached, with probability 1 the walk moves to position 1 at the next step. Otherwise, the walk moves from i to $i+1$ with probability p and from i to $i-1$ with probability $1-p$. Prove that

(a) if $p < 1/2$, each state is positive recurrent.

(b) if $p = 1/2$, each state is null recurrent.

(c) if $p > 1/2$, each state is transient.

Solution to Exercise 7.17: While there are many ways to solve this problem, a particulary pleasant way involves the Catalan numbers; this may require some hints for the students if they have not seen them before. (More on the Catalan numbers can be found in standard texts, or see the excellent Wikipedia article.)

Suppose we start from position k. If we move to $k-1$, it is clear we will eventually (in finite time!) return to k since 0 is a boundary. Hence, it suffices to consider what happens when we move to $k+1$. In this sense, every state has the same behavior with regard to recurrence and transience as state 0, so we just consider starting from state 0 henceforth.

Starting from 0, we must make a move to 1. The probability that our first return to 0 occurs after $2n+2$ moves is given by $C_n p^n (1-p)^{n+1}$, when C_n is the nth Catalan number, given by

$$C_n = \frac{1}{n+1} \binom{2n}{n}.$$

This is because there are C_n paths from 1 back to 1 of length $2n$ that do not hit 0. (Again, this is a standard combinatorial result.)

Recall that we let $r_{0,0}^t$ be the probability that the first return to 0 from 0 is at time t. Then

$$r_{0,0}^t = \sum_{n=0}^{\infty} C_n p^n (1-p)^{n+1}.$$

We now use the Catalan identity

$$\sum_{n=0}^{\infty} C_n x^n = \frac{1 - \sqrt{1 - 4x}}{2x}$$

to find

$$r_{0,0}^t = (1 - p) \sum_{n=0}^{\infty} C_n (p(1-p))^n = (1 - p) \cdot \frac{1 - \sqrt{1 - 4p(1-p)}}{2p(1-p)}.$$

When $p \leq 1/2$, then $\sqrt{1 - 4p(1-p)} = 2p$, and we have $r_{0,0}^t = 1$. Hence when $p \leq 1/2$ the chain is recurrent. When $p > 1/2$, then $\sqrt{1 - 4p(1-p)} = 2 - 2p$, and $r_{0,0}^t = (1-p)/p < 1$. In this case, the chain is transient.

To cope with the issue of whether state 0 is positive recurrent or null recurrent, we consider

$$\begin{aligned} h_{0,0}^t &= \sum_{n=0}^{\infty} (2n+2) C_n p^n (1-p)^{n+1} \\ &= 2(1-p) \sum_{n=0}^{\infty} \binom{2n}{n} (p(1-p))^n. \end{aligned}$$

We make use of the binomial identity

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = \frac{1}{\sqrt{1 - 4x}}$$

to find

$$h_{0,0}^t = \frac{2(1-p)}{\sqrt{1 - 4p(1-p)}}.$$

Hence $h_{0,0}^t$ is finite when $p < 1/2$ and is infinite when $p = 1/2$.

Exercise 7.20: We have considered the gambler's ruin problem in the case where the game is fair. Consider the case where the game is not fair; instead, the probability of losing a dollar each game is $2/3$, and the probability of winning a dollar each game is $1/3$. Suppose that you start with i dollars and finish either when you reach n or lose it all. Let W_t be the amount you have gain after t rounds of play.

(a) Show that $\mathbf{E}[2^{W_{t+1}}] = \mathbf{E}[2^{W_t}]$.

(b) Use part (a) to determine the probability of finishing with 0 dollars and the probability of finishing with n dollars when starting at position i.

(c) Generalize the preceding argument to the case where the probability of losing is $p > 1/2$. (*Hint:* Try considering $\mathbf{E}[c^{W_t}]$ for some constant c.)

Solution to Exercise 7.20:

(a) We have, whenever we not reached 0 or n dollars, that

$$\mathbf{E}[2^{W_{t+1}} \mid W_t] = \frac{2}{3} 2^{W_t - 1} + \frac{1}{3} 2^{W_t + 1} = 2^{W_t}.$$

It follows that

$$\mathbf{E}[2^{W_{t+1}}] = \mathbf{E}[\mathbf{E}[2^{W_{t+1}} \mid W_t]] = \mathbf{E}[2^{W_t}].$$

(b) We have $W_0 = 0$, and inductively, $\mathbf{E}[2^{W_t}] = \mathbf{E}[2^{W_0}] = 1$. Let q_i be the probability that starting from i we eventually end at 0. By the same argument as for the gambler's ruin problem, we now have

$$\lim_{t \to \infty} \mathbf{E}[2^{W_t}] = 1,$$

and hence

$$q_i 2^{-i} + (1 - q_i) 2^{n-i} = 1.$$

We find

$$q_i = \frac{2^n - 2^i}{2^n - 1}.$$

(c) We can repeat the same argument, with $c = p/(1-p)$. We have

$$\mathbf{E}[c^{W_{t+1}} \mid W_t] = p(p/(1-p))^{W_t-1} + (1-p)(p/(1-p))^{W_t+1} = c^{W_t}.$$

We therefore find in this setting

$$q_i c^{-i} + (1-q_i)c^{n-i} = 1,$$

so

$$q_i = \frac{c^n - c^i}{c^n - 1}.$$

Exercise 7.21: Consider a Markov chain on the states $\{0, 1, \ldots, n\}$, where for $i < n$ we have $P_{i,i+1} = 1/2$ and $P_{i,0} = 1/2$. Also, $P_{n,n} = 1/2$ and $P_{n,0} = 1/2$. This process can be viewed as a random walk on a directed graph with vertices $\{0, 1, \ldots, n\}$, where each vertex has two directed edges: one that returns to 0 and one that moves to the vertex with the next higher number (with a self-loop at vertex n). Find the stationary distribution of this chain. (This example shows that random walks on directed graphs are very different than random walks on undirected graphs.)

Solution Exercise 7.21: Clearly π_0, the probability of being at state 0 in the stationary distribution, is just $1/2$, since at each step we move to 0 with probability $1/2$. It follows inductively that $\pi_i = 2^{-i-1}$ for $1 < i < n$, and $\pi_n = 2^{-n}$.

Exercise 7.23: One way of spreading information on a network uses a rumor-spreading paradigm. Suppose that there are n hosts currently on the network. Initially, one host begins with a message. Each round, every host that has the message contacts another host chosen independently and uniformly at random from the other $n - 1$ hosts, and sends that host the message. We would like to know how many rounds are necessary before all hosts have received the message with probability 0.99.

(a) Explain how this problem can be viewed in terms of Markov chains.

(b) Determine a method for computing the probability that j hosts have received the message after round k given that i hosts have received the message after round $k - 1$. (*Hint:* There are various ways of doing this. One approach is to let $P(i, j, c)$ be the probability j hosts have the message after the first c of the i hosts have made their choices in a round; then find a recurrence for P.)

(c) As a computational exercise, write a program to determine the number of rounds for a message starting at one host to reach all other hosts with probability 0.9999 when $n = 128$.

Solution to Exercise 7.23:

(a) There are multiple ways to view this problem in terms of Markov chains. The most natural is to have X_k be the number of hosts that have received the message after k rounds, with $X_0 = 1$. The distribution of X_{k+1} depends only on X_k; the history and which actual hosts have received the message do not affect the distribution of X_{k+1}.

(b) Following the hint, we have $P(i, j, 0) = 1$ if $i = j$ and 0 otherwise. Further

$$P(i, j, c) = P(i, j, c-1)\frac{j-1}{n-1} + P(i, j-1, c-1)\frac{n-j+2}{n-1}.$$

That is, there are two ways to have j hosts with the message after c of the i hosts have made their choices; either j hosts had the message after $c-1$ choices and the cth host chose one of the other $j-1$ hosts with the message already, or $j-1$ hosts had the message after $c-1$ choices and the cth host chose a host without the message.

(c) We use the above reucrrence to track the entire distribution of X_k round by round. A simple program computes the desired values. For $n = 128$, all the hosts have the message with probability (greater than) 0.99 after 17 rounds (if my code is right!). For probability (greater than) 0.9999, it requires 22 rounds.

Notice that even if each round each host contacted a host without the message, 7 rounds would be needed.

Exercise 7.26: Let n equidistant points be marked on a circle. Without loss of generality, we think of the points as being labeled clockwise from 0 to $n - 1$. Initially, a wolf begins at 0, and there is one sheep at each of the remaining $n - 1$ points. The wolf takes a random walk on the circle. For each step, it moves with probability $1/2$ to one neighboring point and with probability $1/2$ to the other neighboring point. At the first visit to a point, the wolf eats a sheep if there is still one there. Which sheep is most likely to be the last eaten?

Solution to Exercise 7.26: Interestingly, each of the remaining sheep is equally likely to be the last eaten. There are many different ways this can be shown. One way is as follows. Consider the sheep at position $n - 1$. This is the last sheep to be eaten if the wolf, starting from 0, walks in one direction from 0 to $n - 2$ before walking in the other direction directly from 0 to $n - 1$. By breaking the circle between $n - 2$ and $n - 1$, we see that his is equivalent to a gambler's ruin problem as in Section 7.2.1. Because of this, the probability the sheep at $n - 1$ is last is $1/((n - 2) + 1) = 1/(n - 1)$. By symmetry the same is true for the sheep at position 1.

Now consider the sheep at position i, $1 < i < n - 1$. There are two ways the sheep could be last: the wolf first walks from 0 to $i - 1$ before reaching $i + 1$ (event A) and then walks from $i - 1$ to $i + 1$ without passing through i (event B); or the wolf first walks from 0 to $i + 1$ before reaching $i - 1$ (event C), and then walks from $i + 1$ to $i - 1$ without passing through i (event D). Again, each of these are simple gambler's ruin problems, and combining them gives

$$\Pr(A) \cdot \Pr(B) + \Pr(C) \cdot \Pr(D) = \frac{i-1}{n-2}\frac{1}{n-1} + \frac{n-i-1}{n-2}\frac{1}{n-1} = \frac{1}{n-1}.$$

Exercise 7.27: Suppose that we are given n records, $R_1, R_2, \ldots, R_n$. The records are kept in some order. The cost of accessing the jth record in the order is j. Thus, if we had 4 records ordered as R_2, R_4, R_3, R_1, the cost of accessing R_4 would be 2, and the cost of accessing R_1 would be 4.

Suppose further that at each step, record R_j is accessed with probability p_j, with each step being independent of other steps. If we knew the values of the p_j in advance, we would keep the R_j in decreasing order with respect to p_j. But if we don't know the p_j in advance, we might use the "move-to-front" heuristic: at each step, put the record that was accessed at the front of the list. We assume that moving the record can be done with no cost and that all other records remain in the same order. For example, if the order was R_2, R_4, R_3, R_1 before R_3 was accessed, then the order at the next step the order would be R_3, R_2, R_4, R_1.

In this setting, the order of the records can be thought of as the state of a Markov chain. Give the stationary distribution of this chain. Also, let X_k be the cost for accessing the kth requested record. Determine an expression for $\lim_{k \to \infty} \mathbf{E}[X_k]$. Your expression should be easily computable in time that is polynomial in n, given the p_j.

Solution to Exercise 7.27: Let us consider an example with 4 records. Let $\pi(i)$ be the record in the ith position. The key is to note that for the records to be in the order $R_{\pi(1)}, R_{\pi(2)}, R_{\pi(3)}, R_{\pi(4)}$, we must have that, thinking of the sequence of choices in *reverse* order, $\pi(1)$ was chosen first, then $\pi(2)$, and so on. This gives the probabilility of this order as

$$\frac{p_{\pi(1)}}{p_{\pi(1)} + p_{\pi(2)} + p_{\pi(3)} + p_{\pi(4)}} \frac{p_{\pi(2)}}{p_{\pi(2)} + p_{\pi(3)} + p_{\pi(4)}} \frac{p_{\pi(3)}}{p_{\pi(3)} + p_{\pi(4)}}.$$

This formula generalizes; if there are n records, the probability is:

$$\prod_{i=1}^{n-1} \frac{p_{\pi(1)}}{\sum_{j=i}^{n} p_{\pi(j)}}.$$

This must correspond to the stationary distribution.

To calculate X_k, consider accessing some record far off in time. In the stationary distribution, record R_j is ahead of record R_i if R_j was chosen more recently than R_i. The probability of this is just $p_j/(p_i + p_j)$. To find $\mathbf{E}[X_k]$, let X_{ij} be the probability that at time k that record i is chosen and record j is ahead of it. Then

$$\mathbf{E}[X_k] = \mathbf{E}\left[\sum_{i,j \neq i} X_{ij}\right] = \sum_{i,j \neq i} \mathbf{E}[X_{i,j}] = \sum_{i,j \neq i} p_i \frac{p_j}{p_i + p_j}.$$

This can be easily calculated in polynomial time.