LECTURE NOTES

CPSC 351 — Winter 2025

Theoretical Foundations of Computer Science II

Philipp Woelfel

Chapter 1: Discrete Probability Theory
Chapter 2: Random Variables

Personal use of this material is permitted. Permission from the author must be obtained for all other uses, including any form of distribution.

Contents

1. Discrete Probability Theory

Modern computer science heavily relies on probability theory. For many algorithmic problems, we can find simpler and or more efficient solutions, if we allow computer programs to make random decisions. Some entire disciplines, such as cryptography, are built on probability theory. Probability theory is also the basis for statistics, and thus fundamental for the design of experiments and interpretation of experimental data.

1.1. The Probability Space

Consider rolling a 6-sided die. What is the probability that the die shows one pip (value 1)? Most people would say it is 1/6. The underlying assumption is that each side of the die occurs with the same probability.

Probability theory defines a mathematical model that can be used to represent what we believe is happening when we perform random experiments, such as rolling a die, flipping a coin, or spinning a roulette ball.

Formally, we consider a random experiment, which is to draw at random an element from a set Ω ("Omega"). The set Ω is called *sample space* (or *event space*), and subsets of that set are called *events*. A single element of Ω is called *elementary event*. Really, these are just new names for set theoretical objects (see also Figure 1.1):

- sample space = universe;
- event = subset of the universe;
- elementary event = element in the universe.

We also need to assign each elementary event $w \in \Omega$ a *probability* $p(w)$. Formally, this is done by defining a function that maps each elementary event to a real number between 0 and 1.

$$p : \Omega \to [0, 1].$$

It is common to denote this function p, Pr, or $Prob$. (In the lectures any of these variants may be used.)

The only property required is that the probabilities of all elementary events sums up to 1:

$$\sum_{w \in \Omega} p(w) = 1.$$

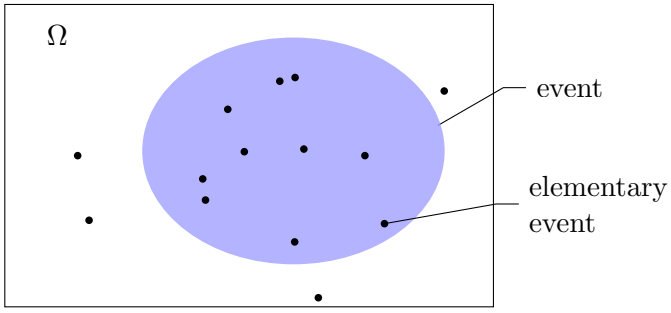

Figure 1.1.: Sample space Ω

We call function p the *probability distribution*. This function is really the only "new" concept, in addition to what we already learned in set theory.

The sample space Ω together with the probability distribution p defines the *probability space*.

To model random events occurring in real life, we only have to define a probability space, i.e., describe the sample space and the probability distribution.

Example 1.1. To model a coin flip we can define the probability space that consists of the sample space $\Omega = \{heads, tails\}$, and the probability distribution $p : \Omega \to [0, 1]$, where $p(heads) = p(tails) = 1/2$. ◀

The model resulting from the definition of a probability space allows us to reason mathematically about probabilities. But whether a probability space (i.e., the model) accurately describes reality is beyond mathematical reasoning. For example, we may assume that when we flip a coin the result will show heads with probability 1/2 and tails with probability 1/2. This assumption leads to the definition of the probability space in Example 1.1. But whether this model accurately describes reality depends on several physical factors, such as the coin, how the coin is flipped, the surface it lands on, etc. A real life coin may land on one side with a slightly higher probability than another, and a skilled magician may be able to flip a coin in such a way that it looks random, but in fact the outcome is predetermined.

Countable Sets

This course is restricted to *discrete* probability theory, which means that we will always assume that the sample space, Ω, is *countable*. Mathematically, a set S is countable, if there exists a surjective (onto) function $f : \mathbb{N} \to S$.

One can prove that a set S is countable is by giving an algorithm that prints all elements of S. Consider the algorithm in Figure 1.2. It prints the integers 0, 0, -1, 1, -2, 2, -3, 3,..., and thus it prints all elements in $\mathbb{Z}$. Therefore, $\mathbb{Z}$ is countable.

Trivially, every finite set is countable. Examples of infinite countable sets are $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and the set of all words using the English alphabet.

An uncountable (i.e., not countable) set is $\mathbb{R}$, the set of real numbers.

It is possible to use uncountable sample spaces, but the math gets a bit more complicated. In particular, wherever we use sums here, we would have to use some form of integrals.

```
i:=0
while True do
    print(−i)
    print(i)
    i := i + 1
end
```

Figure 1.2.: Printing all integers

1.2. The Probability of Events and the Uniform Distribution

The most common assumption is that all outcomes of a random experiment (elementary events) are equally likely. In this case, we say that events are *uniformly distributed*, and we have

$$p(w) = \frac{1}{|\Omega|} \text{ for each } w \in \Omega.$$

The probability distribution used in Example 1.1 is uniform, because it assigned each elementary event in $\Omega = \{heads, tails\}$ the same probability 1/2.

Recall that an *event* is a subset of the sample space Ω. Usually we are interested in the probability of events. For an event $E \subseteq \Omega$ we define the probability of E as the sum of all probabilities of elementary events in E:

$$p(E) = \sum_{w \in E} p(w).$$

Note that if p is the uniform distribution, then this simplifies to:

$$p(E) = \sum_{w \in E} p(w) = \sum_{w \in E} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}.$$

Fact 1.2. *If p is the uniform distribution, then each event E has probability*

$$p(E) = \frac{|E|}{|\Omega|}.$$

Example 1.3. What is the probability that the result of rolling a single die is an even number of pips?

We can model this with a sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$, where each of the events $1, 2, 3, 4, 5, 6$ occurs with the same probability of $1/6$. I.e., $p(w) = 1/6$ for each $w \in \Omega$.

The event that the die shows an even number of pips is $E = \{2, 4, 6\}$. Hence, the probability that this event occurs is

$$p(E) = \frac{|\{2, 4, 6\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6} = \frac{1}{2}. \qquad \blacktriangleleft$$

Example 1.4. Consider an algorithm that outputs a string of 8 random bits.

(a) What is the probability that the output corresponds to the ASCII code of a lowercase letter from the English alphabet?

(b) What is the probability that the bit string contains exactly 3 zeros?

We can model this random experiment with the sample space $\Omega = \{0, 1\}^8$ and the uniform distribution. Since we have $|\Omega| = 2^8 = 256$, we have $p(w) = 1/256$ for each elementary event $w \in \Omega$.

(a) There are 26 lowercase letters in the set $\{a, b, \ldots, z\}$, and each of them is represented by exactly one ASCII code. Let E_1 be the set of the corresponding ASCII codes. Thus, the probability that random bit string corresponds to a lowercase English letter in ASCII is

$$p(E_1) = \frac{|E_1|}{|\Omega|} = \frac{26}{256} = 0.1015625.$$

(b) Let E_2 denote the set of all bit strings of length 8 that contain exactly 3 zeros. Similarly to the example in Lecture 25, we can see that there are $\binom{8}{3}$ such bit strings. Hence,

$$|E_2| = \binom{8}{3} = \frac{8!}{3! \cdot (8-3)!} = \frac{8!}{3! \cdot 5!} = 56.$$

Thus, the probability of getting a bit string with exactly 3 zeros is $56/256 = 0.21875$. $\blacktriangleleft$

Example 1.5. Suppose in a poker game a player receives 5 out of 52 playing cards. What is the probability that 4 of the 5 cards have the same value (e.g., 4 Aces, 4 Kings, etc.)?

We can model using the sample space Ω that consists of all sets of 5 playing cards (out of 52). The probability distribution is uniform, because each such 5 tuple is equally likely to be drawn. Hence, $|\Omega| = \binom{52}{5} = 2\,598\,960$.

Let E be the set of all sets of 5 playing cards, where 4 cards have the same value v. There are 13 values to choose for v, and for each value v there are exactly 4 cards with that value. Hence, there are 13 possibilities to choose the 4 cards of the same value. The 5th card can be anyone among the remaining 48 cards. Hence, we have

$$|E| = 13 \cdot 48 = 624.$$

Thus, the probability of getting 4 cards with the same value in a poker hand of 5 cards, is

$$p(E) = \frac{624}{2\,598\,960} \approx 0.00024 = 0.024\%.$$ ◀

1.2.1. The Probability of the Complement of an Event

Often it is easier to determine the probability of the complement of an event than the probability of the event itself. We can use the following formula.

Theorem 1.6. *Let Ω a sample space, E an event, and $\overline{E} = \Omega \setminus E$. Then*

$$p\big(\overline{E}\big) = 1 - p(E)$$

Proof. Recall that $\sum_{w\in\Omega} p(w) = 1$. Hence,

$$p\big(\overline{E}\big) = \sum_{w\in\Omega\setminus E} p(w) = \sum_{w\in\Omega} p(w) - \sum_{w\in E} p(w) = 1 - \sum_{w\in E} p(w) = 1 - p(E).$$ □

Example 1.7. Consider again a computer program the generates exactly 8 random bits. What is the probability that at least one bit is 1?

We can model this with the sample space $\Omega = \{0,1\}^8$, and the uniform distribution. Let E be the event that at least one bit is 1, i.e., E is the subset of Ω that contains all bit strings that have at least one 1.

It is easier to determine the probability of event $\overline{E} = \Omega \setminus E$. Thus, $\overline{E} = \{00000000\}$, i.e., $\overline{E}$ is the event that all 8 bits are 0.

We have

$$p(E) = 1 - p\big(\overline{E}\big) = 1 - \frac{1}{|\Omega|} = 1 - \frac{1}{2^8} = 1 - \frac{1}{256} = \frac{255}{256}.$$ ◀

1.2.2. The Probability of the Union of Events

Theorem 1.8. *Let A and B be two events. Then*

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$

Proof.

$$p(A \cup B) = \sum_{w \in A \cup B} p(w) = \sum_{w \in A} p(w) + \sum_{w \in B} p(w) - \sum_{w \in A \cap B} p(w) = p(A) + p(B) - p(A \cap B). \quad \square$$

Example 1.9. Suppose we roll a 6-sided and unbiased die twice. What is the probability of seeing at least one 6?

We can model this as follows: The sample space is $\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$. An elementary event $(x_1, x_2) \in \Omega$ describes the result, where the first die roll shows x_1 and the second die roll x_2. The probability distribution is uniform, i.e., each of the 36 outcomes has probability 1/36.

For $i \in \{1, 2\}$ let E_i denote the event that the i-th die roll shows a 6. Then $p(E_i) = 1/6$.

Moreover, $E_1 \cap E_2$ is the event that E_1 occurs **and** E_2 occurs, i.e., both die rolls show a 6. Then $E_1 \cap E_2 = \{(6, 6)\}$, and $p(E_1 \cap E_2) = 1/36$.

Hence, the event $E_1 \cup E_2$ that the first die **or** the second die shows a 6 has probability:

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} \approx 0.3056. \quad \blacktriangleleft$$

Note that Theorem 1.8 implies that if two events A and B are disjoint (i.e., they have an empty intersection), then $p(A \cup B) = p(A) + p(B)$. We can use this in the following example.

Example 1.10. In Canada's lottery Lotto Max 7 distinct numbers out of $\{1, \ldots, 50\}$ are drawn at random (without replacement). A \$5 ticket allows a player to choose three sets (called "lines"), each containing 7 distinct numbers. A Jackpot Win occurs, when one of the player's lines matches the 7 drawn numbers exactly. The Western Canada Lottery Corporation (WCLC) claims that the probability of a Jackpot Win is "$1 : 33,294,800$" (see `https://www.wclc.com/games/lotto-max.htm`). Is this true?

We can model the random experiment of drawing 7 numbers from $\{1, \ldots, 50\}$ without replacement using the sample space $\Omega = \{S \subseteq \{1, \ldots, 50\} \mid |S| = 7\}$ and the uniform distribution. I.e., the sample space Ω contains all subsets of $\{1, \ldots, 50\}$ with exactly 7 elements.

Let E_1, E_2, and E_3 denote the events the player's first, second, and third line, respectively match. Thus, E_i contains exactly one set of 7 numbers from $\{1, \ldots, 50\}$. The probability that one of the player's chosen lines occurs is

$$p(E_1 \cup E_2 \cup E_3).$$

Now, **if the player's lines are all distinct** (i.e., no two of the player's chosen sets of numbers are the same), then by Theorem 1.8 we have

$$p(E_1 \cup E_2 \cup E_3) = p(E_1) + p(E_2) + p(E_3) = 3 \cdot \frac{1}{|\Omega|} = 3 \cdot \frac{1}{\binom{50}{7}} = 3 \cdot \frac{7! \cdot 43!}{50!}.$$

Using a calculator, we can determine that

$$\frac{50!}{3 \cdot 7! \cdot 43!} = 33,294,800.$$

Hence,

$$p(E_1 \cup E_2 \cup E_3) = \frac{1}{33,294,800},$$

exactly as claimed by the WCLC.

But notice that this requires that a player chooses three different lines. If two or all three of the lines are the same, then the player's probability of winning the jackpot is smaller. ◀

1.2.3. The Union Bound

When we analyze randomized algorithms (e.g., their running time or error probability), it is often useful to prove *upper bounds* instead of exact bounds. For example, we want to prove statements like "the running time of an algorithm is *at most*...", or "the probability that the algorithm outputs a wrong answer is *at most*...". The main reason for proving such upper bounds (i.e., inequalities) it that it is often not possible or much harder to obtain exact statements.

Theorem 1.8 immediately yields an *upper bound* on the probability of the union of two events:

$$p(A \cup B) \leq p(A) + p(B).$$

More generally, for multiple events $E_1, \ldots, E_k$ we obtain

$$p(E_1 \cup E_2 \cup \cdots \cup E_k) \leq \sum_{i=1}^{k} p(E_i). \tag{1.1}$$

Note that $E_1 \cup \cdots \cup E_k$ is the event that E_1 occurs or E_2 occurs or ... or E_k occurs.

This inequality is called *union bound*, and it is one of the most useful inequalities for the analysis of randomized algorithms.

Example 1.11 (Three Dice)**.** Suppose we roll three (6-sided) dice. We are interested in the event C that at least two of the three dice end up with the same number of pips on top. We can use the union bound to determine an upper bound for the probability of C.

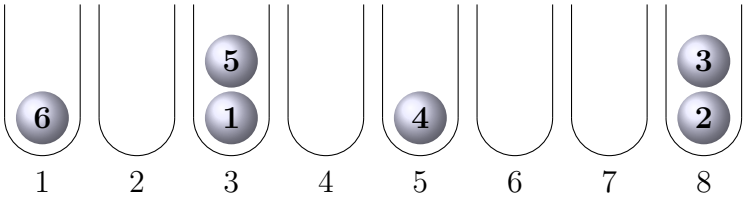

Figure 1.3.: Balls into Bins

The sample space is $\Omega = \{1, \ldots, 6\}^3$, and we use the uniform distribution. Thus, each elementary event is a triple (d_1, d_2, d_3), where d_i is the result of the i-th die. Let $C_{i,j}$ denote the event that dice i and j show the same number. For example, $C_{2,3}$ means that dice two and three show the same number. It is easy to see that $p(C_{i,j}) = 1/6$ as long as $i \neq j$.

Hence, $C = C_{1,2} \cup C_{2,3} \cup C_{1,3}$. By the union bound, the probability of two dice showing the same value can be bounded as follows:

$$p(C) = p(C_{1,2} \cup C_{2,3} \cup C_{1,3}) \leq p(C_{1,2}) + p(C_{2,3}) + p(C_{1,3}) \leq \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Note that this upper bound is not tight, i.e., the actual probability of C is strictly less than $1/2$ (see also Exercise 1.24). ◀

Example 1.12 (Balls into Bins)**.** Suppose we have n balls and m bins. We throw each of the n balls into a randomly chosen bin. We say two balls *collide*, if they are thrown in the same bin. What is an upper bound on the probability that at least two of the n balls collide?

We may assume (w.l.o.g.) that each ball is labelled with a number from $\{1, \ldots, n\}$, and each bin with a number from $\{1, \ldots, m\}$. Let $E_{i,j}$ denote the event that ball i and ball j end up in the same bin. For distinct balls i and j, we have

$$p(E_{i,j}) = \frac{1}{m}.$$

There are two ways to argue that: First, suppose ball i ends up in some bin, say b. The probability that ball j ends up in the same bin is $1/m$. (See also Section 1.3.5 for an explanation why we can argue this way.)

Another way of seeing that is by counting as follows: There are m^2 of possible pairs (b_i, b_j) of bins, such that ball i chooses bin b_i and ball j chooses bin b_j. There are m such pairs that correspond to balls i and j being in the same bin, namely the pairs $(1,1), (2,2), \ldots, (m,m)$. Hence, the probability that balls i and j collide is $m/m^2 = 1/m$.

Two of the balls $1, \ldots, n$ collide if and only if there exists $i, j \in \{1, \ldots, n\}$ such that $E_{i,j}$ occurs. Note that $E_{i,j}$ is the same as $E_{j,i}$, so we only need to consider events $E_{i,j}$ with $1 \leq i < j \leq n$.

Hence, we are interested in the event

$$E = \bigcup_{1 \le i < j \le n} E_{i,j} = E_{i,j}.$$

Note that there are $\binom{n}{2}$ pairs (i, j) with $i < j$. Using the union bound we obtain the following upper bound for the probability that at least two balls collide:

$$p(E) \le \sum_{1 \le i < j \le n} p(E_{i,j}) = \sum_{1 \le i < j \le n} \frac{1}{m} = \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2m} < \frac{n^2}{2m}.$$

For example, if we choose $m = n^2$, then the probability that we have a collision is less than $1/2$. ◀

The example above has an important algorithmic application: Suppose we want to hash n keys into an array of size m. I.e., we are given a set S of n keys from some universe U, and a hash function $h : U \to \{0, \ldots, m-1\}$. The goal is to store each key $x \in S$ in an array $A[0..m-1]$ in the array entry $A[y]$, where $y = h(x)$. This is only possible if there are no *collisions*, i.e, no two distinct keys $x, x' \in S$ are hashed to the same array position $h(x) = h(x')$. (Of course we can also deal with collisions using more complex data structures, such as linked lists in hashing with chaining.)

Under the assumption that the sequence of n hash function values of the n keys is uniformly distributed (over the set $\{0, \ldots, m-1\}^n$), this corresponds exactly to the balls-into-bins problems described above: The i-th key corresponds to the i-th ball, and the j-th array entry to the j-th bin.

1.3. Conditional Probabilities and Independence

See Section 7.2.4 in the textbook.

Definition 1.13. *Let A and B be events, where $p(B) > 0$. The* conditional probability *of A* given *B is denoted $p(A \mid B)$, and defined as*

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}.$$

Note that $p(A \mid B)$ is not defined, if $p(B) = 0$.

Example 1.14 (Black Jack)**.** In Black Jack a player is initially dealt two cards. Each card has a value, and the goal is to get a sum of values as close to 21 as possible, but without exceeding the value of 21. A player is initially dealt two cards, and is then allowed to ask the dealer to deal her additional cards, as long as her current hand has a total value of less than 21.

Suppose we play Black Jack with our friend, who holds two cards drawn at random from a deck of 52 in her hand. She tells us that exactly one of the two cards has a value of 10. This means that this card can be any one of 10, jack, queen, and king. What is the probability that the other card is an ace (which has a value of 11)?

Let V be the set of 52 cards. The sample space is $\Omega = \{S \in \mathcal{P}(V) \mid |S| = 2\}$.

Let B be the event "exactly one of the two cards is one of 10, jack, queen, and king". Thus, B is the set of all sets $\{x, y\}$ that contain exactly one of those cards.

There are 16 choices for those cards, and there are 52-16=36 choices for the other card. Thus, $|B| = 16 \cdot 36$, and so

$$p(B) = \frac{16 \cdot 36}{|\Omega|}.$$

Let A be the event that "one card is an ace". Then $A \cap B$ is the event that "one card is an ace, and the other card is one of 10, jack, queen, and king". Since there are 4 choices for the ace and 16 choices for the other card being 10, jack, queen, or king, we have $|A \cap B| = 4 \cdot 16$. Hence,

$$p(A \cap B) = \frac{4 \cdot 16}{|\Omega|}.$$

Therefore, the probability that one card is an ace, if exactly one card is one of 10, jack, queen, and king is:

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{4 \cdot 16/|\Omega|}{16 \cdot 36/|\Omega|} = \frac{1}{9}.$$ ◀

1.3.1. The Conditional Probability Space

If C is an event in Ω, then all probability rules apply for conditional events "given C", as for unconditional events. For example, the rule $p(A) = 1 - p\big(\overline{A}\big)$ translates to

$$p(A \mid C) = 1 - p\big(\overline{A} \mid C\big).$$

Similarly, the inclusion-exclusion principle, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$ implies

$$p(A \cup B \mid C) = p(A \mid C) + p(B \mid C) - p(A \cap B \mid C).$$

More generally, C defines a new *conditional* probability space *given* event C. It uses the same sample space Ω, and the probability distribution is $p' : \Omega \to [0, 1]$, where

$$p'(A) = p(A \mid C).$$

1.3.2. The Law of Total Probability

The following theorem describes the law of total probability.

Theorem 1.15. *For any two events A and B:*

$$p(A) = p(A \mid B) \cdot p(B) + p\big(A \mid \overline{B}\big) \cdot p\big(\overline{B}\big).$$

Example 1.16. Suppose we are given two piles of cards. The cards on one pile has only blue backs, the ones on the other pile have red backs. The blue back pile contains an Ace and a King, the red back pile contains an Ace, a King, and a Queen.

- Blue back: Ace, King.
- Red back: Ace, King, Queen.

We first choose a random pile, the blue one or the red one, each with equal probability. Then we choose one of the cards in that pile at random. What is the probability that the chosen card is an Ace?

Let R be the event that we choose the red pile (and thus $\overline{R}$ is the event that we choose the blue pile). Further, let A be the event that we select the Ace from the chosen pile. Then the probability of choosing an Ace is:

$$p(A) = p(A \mid R) \cdot p(R) + p\big(A \mid \overline{R}\big) \cdot p\big(\overline{R}\big) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{6} + \frac{1}{4} = \frac{2}{12} + \frac{3}{12} = \frac{5}{12}. \quad \blacktriangleleft$$

The total probability theorem can be generalized as follows:

Theorem 1.17. *Let A be an event, and let $B_1, \ldots, B_k$ be disjoint events, such that $B_1 \cup \cdots \cup B_k = \Omega$. Then*

$$p(A) = p(A \mid B_1) \cdot p(B_1) + p(A \mid B_2) \cdot p(B_2) + \cdots + p(A \mid B_k) \cdot p(B_k).$$

1.3.3. Bayes' Theorem

Solving the equation in Definition 1.13 for $p(A \cap B)$ and $p(B \cap A)$, leads to a useful identity:

$$p(A \cap B) = p(A \mid B) \cdot p(B) = p(B \mid A) \cdot p(A). \tag{1.2}$$

An equivalent statement is the following:

Theorem 1.18 (Bayes' Theorem)**.**

$$p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)}.$$

Example 1.19. Assume the following:

1. 1 in 100 people has cancer.
2. A cancer screening test has a sensitivity of 80%, meaning that 80 out of 100 people who have cancer get a positive test result. I.e., the false negative rate is 20%.
3. The same test has a specificity of 90%, meaning that 90 out of 100 people who do not have cancer get a negative test result. I.e., the false positive rate is 10%.

Suppose a person takes a cancer screening test, and it comes back positive. What is the probability that that person has indeed cancer?

Let C be the event that a person has cancer, and P the event that a screening test is positive. The above assumptions can be expressed mathematical as follows:

1. $p(C) = 0.01$.
2. $p(P \mid C) = 0.8$.
3. $p\big(\overline{P} \mid \overline{C}\big) = 0.9$.

We are interested in the conditional probability of C (the person has cancer), given P (the screening test is positive). According to Bayes' Theorem, this is

$$p(C \mid P) = \frac{p(P \mid C)p(C)}{p(P)} \tag{1.3}$$

We already know $p(C)$ and $p(P \mid C)$. Thus, we can compute $p(P)$ using the Total Probability Theorem:

$$p(P) = p(P \mid C)p(C) + p\big(P \mid \overline{C}\big)p\big(\overline{C}\big) = 0.8 \cdot 0.01 + (1 - 0.9) \cdot (1 - 0.01) = 0.107.$$

Plugging this into eq. (1.3), we get

$$p(C \mid P) = \frac{0.8 \cdot 0.01}{0.107} = 0.074\ldots$$

Thus, even with a positive screening test result, the probability of having cancer is less than 7.5%. ◀

1.3.4. Independence

Let A and B be events. We say A and B are *independent*, if

$$p(A \cap B) = p(A) \cdot p(B).$$

Observe that if the above identity is true, then

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} = \frac{p(A) \cdot p(B)}{p(B)} = p(A).$$

Hence, the conditional probability that A occurs given B is the same as the unconditional probability of A occurring. This matches our informal understanding of even A not depending on event B.

Example 1.20. Suppose we roll a red die and a blue die. Let A be the probability that the red one shows a 6, and B the event that the blue one shows a 6. Then $p(A) = p(B) = 1/6$, and $p(A \cap B) = 1/36 = p(A) \cdot p(B)$. Hence, events A and B are independent. ◀

Example 1.21. Suppose we roll two dice, one is red the other blue. Let A be the event that the red die comes up 6, and B the event that the sum of pips of both dice is even.

We use the sample space $\Omega = \{1, \ldots, 6\}^2$. Thus, each elementary event is a pair (x, y), where x indicates the value of the red die and y the value of the blue die.

Clearly, $p(A) = 1/6$ and $p(B) = 1/2$. Moreover, $A \cap B$ contains the events $\{(6, 2), (6, 4), (6, 6)\}$, and thus $p(A \cap B) = 3/36 = 1/12$. It follows that $p(A \cap B) = p(A) \cdot p(B)$, so the events A and B are independent. ◀

Example 1.22. Now consider the same setup as in Example 1.21, but assume that the blue die shows only 1, 2, or 3 pips, and each number of pips is shown on two of the 6 sides. The sample space is now $\Omega = \{1, \ldots, 6\} \times \{1, 2, 3\}$, and thus $|\Omega| = 18$. As before, let A be the event that the red die shows 6 pips, and B the event that the sum of pips is even.

Thus, $p(A) = 1/6$. Moreover, $B = \{(1, 1), (1, 3), (2, 2), (3, 1), (3, 3), (4, 2), (5, 1), (5, 3), (6, 2)\}$. Thus, $p(B) = 9/18 = 1/2$. Finally, $A \cap B = \{(6, 2)\}$, so $p(A \cap B) = 1/18 \neq p(A) \cdot p(B)$.

Hence, events A and B are not independent. ◀

1.3.5. Fixing Partial Random Choices

In this section we introduce a technique that can sometimes be used to simplify the computation of probabilities. It can informally be described as follows:

> *Suppose we execute a random experiment, and are interested in the probability of some event A. If a partial result of the random experiment can be fixed in any way without changing the (conditional) probability of A occurring, then we may arbitrarily fix that partial outcome to determine the probability of A.*

Mathematically, this can be described as follows:

Observation 1.23. *Let Ω be a sample space, A an event, and let $B_1 \ldots, B_k$ be disjoint events such that $B_1 \cup \cdots \cup B_k = \Omega$ and $q = p(A \mid B_1) = p(A \mid B_2) = \cdots = p(A \mid B_k)$. Then $p(A) = q$. (Equivalently, A and B_i are independent, for each $i \in \{1, \ldots, k\}$.)*

Proof. By Theorem 1.15:

$$\begin{aligned} p(A) &= p(A \mid B_1) \cdot p(B_1) + \cdots + p(A \mid B_k) \cdot p(B_k) \\ &= q \cdot p(B_1) + \cdots + q \cdot p(B_k) \\ &= q \cdot \big(p(B_1) + \cdots + p(B_k)\big) \\ &= q \cdot 1. \end{aligned}$$

□

Example 1.24. Suppose we roll two dice, a red one and a blue one. What is the probability of event A: "both dice show the same number of pips"?

For $i \in \{1, \ldots, 6\}$ Let R_i be the event that the red die shows i pips. Then $p(A \mid R_i) = 1/6$, because this is the probability that the blue die also shows i pips. By Observation 1.23, $p(A) = 1/6$.

In other words: We can fix the result of the red die arbitrarily, and no matter how we fix it, the probability that both dice show the same number of bits will always be $1/6$. Therefore, the probability that both dice show the same result is $1/6$. ◀

Example 1.25. Consider the following random experiment: We choose 100 random bits $b_1, b_2, \ldots, b_{100} \in \{0, 1\}$ uniformly and independently at random. What is the probability of event E: "$b_1 + b_2 + \cdots + b_{100}$ is even"?

It seems rather obvious that the probability is $1/2$. But what is a rigorous argument? A rather tedious way is to compute the number of combinations of bit strings whose sum is 0, and divide it by the total number of bit strings of length 100.

Instead, we *fix* the first 99 bits, $b_1, \ldots, b_{99}$ arbitrarily. We then observe that no matter how we fix the first 99 bits, it does not affect the probability of event E: If the sum of the first 99 bits is

even, then E occurs if and only if bit b_{100} is also even, and thus with probability $1/2$. If the sum of the first 99 bits is odd, then E occurs if and only if bit b_{100} is also odd, and thus again with probability $1/2$. Hence, no matter how we fix the first 99 bits, the probability of E occurring is exactly $1/2$, and so $p(E) = 1/2$. ◀

1.3.6. Mutual Independence

See Definition 5 in the textbook.

If multiple events $E_1, E_2, \ldots, E_n$ are mutually independent, then

$$p(E_1 \cap E_2 \cap \cdots \cap E_n) = p(E_1) \cdot p(E_2) \cdots p(E_n). \tag{1.4}$$

(Note that the opposite is not true, i.e., if $p(E_1) \cdot p(E_2) \cdots p(E_n) = p(E_1 \cap E_2 \cap \cdots \cap E_n)$, then this does not necessarily imply that the events are mutually independent.)

This is very useful to determine the probability of events. For example, suppose we throw n balls into m bins, and the choices of bins are (mutually) independent for all balls. What is the probability that all balls land in the same bin?

Let A_j for $j \in \{1, \ldots, m\}$ be the event that all balls land in bin j. We will first determine $p(A_1)$, i.e., the probability that all balls land in bin 1.

For $i \in \{1, \ldots, n\}$, let B_i be the event that ball i lands in bin 1. Then $p(B_i) = 1/m$. Since the bin choices are independent for all balls, events $B_1, \ldots, B_n$ are independent. Hence, we obtain from eq. (1.4)

$$p(A_1) = p(B_1 \cap B_2 \cap \cdots \cap B_n) = p(B_1) \cdot p(B_2) \cdots p(B_n) = \left(\frac{1}{m}\right)^n.$$

Obviously, $p(A_1) = p(A_2) = \cdots = p(A_m)$. Moreover, all events $A_1, \ldots, A_m$ are disjoint, because it is not possible that A_j and $A_{j'}$ occur at the same time for two distinct bins j and j' (i.e., it is not possible that all balls land in bin j and all balls land in bin j'). Hence, the probability that all balls land in the same bin is

$$\begin{aligned} p(A_1 \cup A_2 \cup \cdots \cup A_m) &= p(A_1) + p(A_2) + \cdots + p(A_m) \\ &= \sum_{j=1}^{m} p(A_j) = \sum_{j=1}^{m} \left(\frac{1}{m}\right)^n = m \cdot \left(\frac{1}{m}\right)^n = \frac{1}{m^{n-1}}. \end{aligned}$$

1.4. Infinite Sample Spaces

All previous examples used a finite sample space Ω. But often, the probability space is infinite. *Discrete* probability theory considers only *countable* sets Ω—the math for uncountable sets gets considerably harder. Most problems we encounter in Computer Science can be modelled with countable probability spaces.

Consider the following random experiment: Alice and Bob take turns flipping a coin. Whoever gets heads first, wins. What is the probability that Alice wins?

We are interested in the number of coin flips until the first heads appears. Thus, we can model the sample space as the set of all strings that have a number of Ts followed by exactly one H. I.e.,

$$\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\} = \big\{T^k H \;\big|\; k \in \mathbb{N}\big\}.$$

The probability function is defined as follows:

$$p\big(T^k H\big) = \frac{1}{2^k} \cdot \frac{1}{2} = \frac{1}{2^{k+1}}.$$

The reason is that we get consecutive k's probability $1/2^k$, and then heads with probability $1/2$.

To check that this is indeed a probability space, we need to make sure that the sum of all probabilities of elementary events sums up to 1:

$$\sum_{k\in\mathbb{N}} p\big(T^k H\big) = \sum_{k\in\mathbb{N}} \frac{1}{2^{k+1}} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1.$$

Suppose Alice flips the first coin. Then she wins, if the result is H, TTH, $TTTTH$, and so on. I.e., if there is an even number of tails followed by heads:

$$p(\text{Alice wins}) = p\big(\big\{T^{2i} H \;\big|\; i \in \mathbb{N}\big\}\big) = \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \cdots = \sum_{i=0}^{\infty} \frac{1}{2^{2i+1}}$$
$$= \frac{1}{2} \cdot \sum_{i=0}^{\infty} \frac{1}{4^i} = \frac{1}{2} \cdot \frac{1}{1 - 1/4} = \frac{2}{3}.$$

Remark: Geometric Series

In the above calculation, we used a well known identity for infinite sums, called *geometric series*: For any real number $x \neq 1$ it holds

$$\sum_{i=0}^{n} x^i = \frac{1 - x^{n+1}}{1 - x}.$$

Calculating the limit for n approaching infinity, we obtain

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1 - x} \quad \text{if } |x| < 1.$$

1.5. Exercises

probability space; probabilities

1.1 Suppose we roll two 12-sided dice, where each die has for each number $i \in \{1, \ldots, 12\}$ exactly one side with i pips. One of the two dice is red, the other one is blue.

(a) Define a sample space and probability function that models rolling both dice and reading off the number of pips each die shows.

(b) What is the probability that the number of pips shown is exactly 17?

(c) For which value of D do we achieve the largest probability that the sum of pips is exactly D?

poker; probabilities

1.2 A poker hand is a selection of 5 distinct cards from a standard deck of 52 cards. Describe a probability space for the random experiment of randomly choosing a poker hand. Then determine the probabilities of the following events (see `https://en.wikipedia.org/wiki/List_of_poker_hands` for a list of poker hands):

(a) The poker hand does not contain the Queen of Diamonds.

(b) At least one card in the poker hand is an Ace.

(c) The poker hand contains two pairs (but not three of a kind).

(d) The poker hand contains a full house.

(e) The poker hand contains a flush.

lotto; probabilities

1.3 The WCLC website `https://www.wclc.com/games/lotto-max.htm` lists the probabilities of various lotto results. Determine the probabilities of the following events, and compare your result to those listed on the website:

(a) 7 of 7

(b) 6 of 7 plus bonus

(c) 6 of 7 (this event occurs only if the bonus number does not match)

1.4 Suppose we roll 3 fair dice. Let A be the event that at least two of the 3 dice show the same number of pips. Determine $p(A)$ exactly. Compare your result with the upper bound proved in Example 1.11.

1.5 The following is taken from Wikipedia (`https://en.wikipedia.org/wiki/Monty_Hall_problem`):

Monty Hall; basic probabilities

> *Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?*

Assume that the car is behind a random door. Further, assume that the host *always* opens a door behind which is a goat, no matter what your choice is. Determine the probabilities of winning the car if you switch, and of winning the car if you don't switch.

1.6 Assume that 23 people meet in a room, who were all born in the same leap year, and whose birthdays are distributed uniformly over 366 days. Show that the probability that at least two of them have the same birthday is greater than 1/2.

birthday paradox; conditional probabilities

1.7 Consider the same setting as in Exercise 1.6.

union bound

(a) Use the union bound to determine an upper bound for the probability that at least one person's birthday is January 1st.

(b) Use the union bound to prove that the probability that at least two of the people in the room have the same birthday is less than 0.7.

Hint: Consider all events, in which two people have the same birthday.

1.8 A common problem in computer science is to store each element from a given set S in a unique entry of some array. For example, S could be the keys of records that have to be stored in a database.

hashing; independence; probabilities

Suppose S is a subset from some universe U. If we have an array of size m, we can use a *hash function* $h : U \to \{0, \ldots, m-1\}$. A *collision* occurs when to elements from S are mapped to the same array position, i.e., if there are two distinct keys $x, x' \in S$ with $h(x) = h(x')$. Provided that no collisions occur, each element $x \in S$ can be stored in the array at position $h(x)$.

Assume that the hash function values $h(x)$ for all elements in S are completely random in $\{0, \ldots, m-1\}$ (i.e., they are independent and uniformly

distributed). What is the probability that a collision occurs, if there are 10 keys (i.e., $|S| = 10$) and

(a) $m = 10$.

(b) $m = 100$.

1.9 Suppose we throw n balls with distinct labels in $\{1, \ldots, n\}$ into m bins with distinct labels in $\{1, \ldots, m\}$. For each ball we choose a bin uniformly and independently at random.

union bound

Use the union bound to derive upper bounds for the probabilities of the following events:

(a) At least 2 balls land in bin 1.

(b) At least one bin remains empty.

(c) There is a bin that contains at least k balls.

As a challenge you may also try to determine exact probabilities for the above events, and reason about the difficulties you may encounter.

1.10 Suppose 50% of all rainy days start off cloudy. Even though 40% of all days start cloudy, it rains only on 10% of all days.

Bayes' Theorem

You wake up and notice that the morning is cloudy. What is the probability that it rains?

1.11 A glazier buys his glass from four different manufacturers - Clearglass (10%), Strongpane (25%), Mirrorglass (30%) and Reflection (35%). In the past, the glazier has found that 1% of Clearglass' product is cracked, 1.5% of Strongpane's product is cracked, and 2% of Mirrorglass' and Reflection's products are cracked.

Bayes' Theorem; Total Probability Theorem

The glazier removes the protective covering from a sheet of glass without looking at the manufacturer's name—in other words, it's a random choice. He finds the glass is cracked. What is the probability it was made by Mirrorglass?[1]

1.12 Suppose A and B are events such that $p(A) = 0.7$ and $p(B) = 0.5$. Prove the following inequalities:

probability inequalities

(a) $p(A \cup B) \geq 0.7$; and

[1]This question is taken from Mathopolis (`https://www.mathopolis.com/questions/q.html?id=11282`.)

(b) $p(A \cap B) \geq 0.2$.

1.13 Consider two dice, one red and one blue. Each die has 6 sides with the numbers 1,2,...,6. But both dice are loaded: For the red die, the probability that a 3 occurs is twice as high as the probability of any other number, and the outcomes 1, 2, 4, 5, and 6 are equally likely. Similarly, for the blue die the probability that a 4 occurs is twice as high as the probability of any other number, and the outcomes 1, 2, 3, 5, and 6 are equally likely.

probability space; non-uniform distribution

(a) Model the probability space for throwing the red die, i.e., describe the sample space and probability distribution p by giving $p(w)$ for each elementary event w.

(b) What is the probability distribution for the blue die?

(c) Model the probability space for rolling the red and the blue die at the same time, and determine the probability that the sum of the pips is 7.

1.14 Solve the following questions using the definition of conditional probabilities: $p(A \mid B) = p(A \cap B)/p(B)$.

conditional probabilities

(a) Suppose on parking lot 11 there are 400 vehicles. 120 of them are red, 60 are trucks, and 20 are red trucks. If a randomly chosen red vehicle leaves the parking lot, what is the probability that it is a truck?

(b) Consider a deck of cards, where the backs are marked in such a way that you can distinguish a pip card (Ace, 2,...,10) from a face card (King, Queen, or Jack). You deal a random poker hand (5 out of 52 cards) to your friend. Your friend holds the cards in such a way that you can only see their backs. Since the backs are marked, you can tell that your friend has three pip cards and two face cards. What is the probability that your friend holds a full house?

1.15 Consider a random bit string of length n. All bits are chosen independently at random, but the probability that a bit is 1 is .51, and the probability that it is 0 is .49. Determine the probabilities of the following events:

bit strings; independence; non-uniform distribution

(a) there are exactly three 1s;

(b) at least one bit is 1;

(c) all bits are the same.

1.16 Decide if in the following events A and B are independent:

independent or not

(a) A: All three die rolls yield the same result; B: the sum of all die roll results is even.

(b) A: The first die roll yields a one; B: the second die roll yields a six.

1.17 Decide if the following events A and B are independent: independent or not

(a) A is the event that a random poker hand (5 out of 52) contains an Ace, and B is the event that it contains a King.

(b) A is the event that a die roll shows a 1 and B is the event that it shows a 6.

(c) A is the event that if a blue and a red die is rolled, the blue one shows a 6, and B is the event that the red one shows a 6.

(d) For the random experiment described in Exercise 1.9, and for $2 \leq n \leq m$, let A be the event that balls 1 and 2 collide, and B the event that balls 2 and 3 collide.

(e) Now let A be the event that balls 1, 2, and 3 collide, and B the event that balls 2, 3, and 4 collide.

1.18 Two players play a game with dice. Whenever it is a player's turn, that player gets to roll one die 3 times. If one of the 3 die rolls shows a 6, that player wins. The players keep taking turns until someone wins the game. infinite probability space

(a) What is the probability that when one player rolls a die 3 times, at least one of the die rolls shows a 6?

(b) Model the probability space of the game described above.

(c) What is the probability that the first player wins? (*Hint:* To calculate the probability use the Geometric Series on p. 17).

(d) How does the answer change, if there are 3 players?

1.19 Consider a finite S of at least n random integers. Choose n integers $s_1, \ldots, s_n$ without replacement from S. I.e., $s_1, \ldots, s_n$ are all distinct, and each s_i, $i \in \{1, \ldots, n\}$, is uniformly distributed over S. Given some $i \in \{1, \ldots, n\}$, what is the probability that s_i is larger than each of $s_1, \ldots, s_{i-1}$? To find a simple solution, partially fix some random choices made. fixing partial random choices

1.20 Suppose A and B are independent events. For each of the following statements, decide if it is true (and give a proof) or false (and give a counter example): independence

(a) $\overline{A}$ and B are independent.

(b) $\overline{A}$ and $\overline{B}$ are independent.

(c) $A \cup B$ and A are independent.

countability

1.21 Recall that a set S is countable, if there is an algorithm that prints all elements in S (see also p. 3). Prove that the set $\mathbb{Q}_{>0}$ of all positive rational numbers is countable.

Hint: You can use two nested while-loops, to print the fractions a/b for all positive integers a and b. But you have to make sure that the inner while-loop terminates for every iteration of the outer while-loop...

applications, probabilities, prime numbers

1.22 Consider two servers, each of which stores 1 TB of data. Note that 1 TB equals 1000^4 bytes, which equals $8 \cdot 1000^4$ bits. Theses two pieces of data can be interpreted as integers x, y, where $0 \leq x, y < 2^{8 \cdot 1000^4}$. We want to decide if x and y are equal using the following random test:

The first server chooses a random 8 byte **prime** number p, i.e., $2 \leq p < 2^{64}$. Then it computes $x' = x \bmod p$ and sends x' and p to the second server. The second server computes $y' = y \bmod p$, and compares the result with x'. If $x' = y'$, the second server outputs "yes", and otherwise it outputs "no".

(a) What is the probability of a false negative, i.e., the server outputs "no", even though $x = y$?

(b) Show that for any positive integer d there are at most $\log_2(d)$ primes that can divide d.

(c) Using (b), show that the probability of a false positive is very small. For this part you can assume without proof the following: "For every positive integer $k > 67$, there are more than $k/\log(k)$ primes in $\{2, \ldots, k\}$."

challenge question

1.23 **Challenge Question:** One hundred people line up to board an airplane, but the first has lost his boarding pass and takes a random seat instead. Each subsequent passenger takes his or her assigned seat if available, otherwise a random unoccupied seat. What is the probability that the last passenger gets his own seat?

Union bound vs. exact bound

1.24 Suppose we roll 3 dice. Determine the exact probability that at least two of them show the same number of pips.

1.25 Assume a family adopts three random pets. Each chosen pet is with probability 1/2 a cat and with probability 1/2 a dog. Let A be the event that the family adopts at least one cat and at least one dog. Further, let B be the event that the family adopts at most one cat. Are events A and B independent?

Independent or not

1.6. Selected Solutions

Exercise 1.5 Recall that our first choice is door 1. Then we are shown a door with a goat behind, and we are asked if we want to switch.

If we don't switch, the probability of winning the car is exactly 1/3, because the car is behind door 1 with probability 1/3. (No matter where the car is, the host can always show us a door with a goat, and the host's information does not affect where the car is.)

Now suppose we do switch. The host shows us a door with a goat. If the car is behind door 1, we switch to the other door that has a goat, and so we lose. This happens with probability 1/3.

If the car is behind door 2, the host will show us door 3, and so we switch to door 2. Thus, we win. This also happens with probability 1/3.

Finally, if the car is behind door 3, the host will show us door 2, and we switch to door 3. Again, we win, and again this happens with probability 1/3.

Hence, if the car is behind door 2 or 3 we win, and if it is behind door 1, we lose. Thus, the probability of winning is 2/3.

Exercise 1.6 The solution can be found in Section 7.2.8 in the textbook.

Exercise 1.7 (b) For any two persons i, j with $1 \leq i < j \leq 23$, let $B_{i,j}$ be the event that persons i and j have the same birthday. Then $p(B_{i,j}) = 1/366$ (fix one person's birthday, and the probability that the other person has the same birthday is always 1/366). Thus, the probability that there are two people with the same birthday is

$$p\left(\bigcup_{1\leq i<j\leq 23} B_{i,j}\right) \leq \sum_{1\leq i<j\leq 23} \frac{1}{366} = \frac{\binom{23}{2}}{366} < 0.7.$$

Exercise 1.9

(a) Let A be the event that at least 2 balls land in bin 1. Further, let B_i be the event that ball i lands in bin 1, for $i \in \{1, \dots, n\}$. Then for any $1 \leq i < j \leq n$, we have $p(B_i \cap B_j) = 1/m^2$. Thus, we obtain

$$p(A) = p\left(\bigcup_{1 \leq i < j \leq n} (B_i \cap B_j)\right) \leq \sum_{1 \leq i < j \leq n} p(B_i \cap B_j) = \sum_{1 \leq i < j \leq n} \frac{1}{m^2} = \frac{\binom{n}{2}}{m^2}.$$

(b) Let E denote the event that at least one bin remains empty. Let E_i denote the event that bin i remains empty, for $i \in \{1, \dots, m\}$. Thus, $E = E_1 \cup \dots \cup E_m$.

Further, let $B_{i,j}$ be the even that ball j lands in bin i for $j \in \{1, \dots, n\}$. Then $p(B_{i,j}) = 1/m$, and so $p\big(\overline{B_{i,j}}\big) = 1 - 1/m = (m-1)/m$. Moreover, since all events $B_{i,1}, \dots, B_{i,n}$ are independent (because each ball chooses its bin independently of other balls):

$$p(E_i) = p\big(\overline{B_{i,1}} \cap \overline{B_{i,2}} \cap \dots \cap \overline{B_{i,n}}\big) = p\big(\overline{B_{i,1}}\big) \cdot p\big(\overline{B_{i,2}}\big) \cdot \dots \cdot p\big(\overline{B_{i,n}}\big) = (1 - 1/m)^n.$$

Finally, by the union bound

$$p(E) = p(E_1 \cup \dots \cup E_m) \leq p(E_1) + \dots + p(E_m) = m \cdot (1 - 1/m)^n.$$

(c) Fix some bin i. First we want to bound the probability that at least k balls land in bin i.

Consider a set $S \subseteq \{1, \dots, n\}$ of size exactly k. Let B_S be the event that all balls in S (and possibly other balls) lands in bin i. Suppose $S = \{b_1, b_2, \dots, b_k\}$. Then

$$\begin{aligned} p(B_S) =& p(b_1 \text{ in } 1 \wedge b_2 \text{ in } 1 \wedge \dots \wedge b_k \text{ in } 1) \\ \overset{\text{ind.}}{=}& p(b_1 \text{ in } 1) \cdot p(b_2 \text{ in } 1) \cdots p(b_k \text{ in } 1) = \left(\frac{1}{m}\right)^k, \end{aligned}$$

where the identity marked "ind." is true because the bins for all balls are chosen independently.

Now let A_i be the event that bin i contains at least k balls. If this is the case, then there must exist a set $S \subseteq \{1, \dots, n\}$, where $|S| = k$, such that bin 1 contains all balls in S. Hence,

$$|A_i| = \bigcup_{\substack{S \subseteq \{1, \dots, n\} \\ |S| = k}} B_S.$$

Thus, we can apply the union bound ("u.b.") as follows:

$$p(A_i) = p\left(\bigcup_{\substack{S \subseteq \{1, \dots, n\} \\ |S| = k}} B_S\right) \overset{\text{u.b.}}{\leq} \sum_{\substack{S \subseteq \{1, \dots, n\} \\ |S| = k}} p(B_S) = \sum_{\substack{S \subseteq \{1, \dots, n\} \\ |S| = k}} \left(\frac{1}{m}\right)^k.$$

The number of terms in this sum is equal to the number of subsets of $\{1, \ldots, n\}$ of size k. Since there are $\binom{n}{k}$ such subsets, we have

$$p(A_i) \leq \binom{n}{k}\left(\frac{1}{m}\right)^k.$$

Now we can apply the union bound again, to bound the probability that there exists a bin with at least k balls:

$$p(A_1 \cup \cdots \cup A_m) \leq p(A_1) + \cdots + p(A_m) \overset{\text{u.b.}}{\leq} m \cdot \binom{n}{k}\left(\frac{1}{m}\right)^k = \binom{n}{k}\left(\frac{1}{m}\right)^{k-1}.$$

Exercise 1.10 Let R denote the event that a day is rainy, and C that a day starts off cloudy. Then we have the following probabilities:

- $p(C \mid R) = 0.5$ (50% of all rainy days start off cloudy).
- $p(C) = 0.4$ (40% of all days start off cloudy).
- $p(R) = 0.1$ (it rains on 10% of all days).

We want to determine the conditional probability that it rains, given that a day starts cloudy, i.e., $p(R \mid C)$.

By definition of conditional probabilities (or Bayes' Theorem):

$$p(R \mid C) = \frac{p(R \cap C)}{p(C)} = \frac{p(C \mid R)p(R)}{p(C)} = \frac{0.5 \cdot 0.1}{0.4} = 0.125.$$

Exercise 1.13

(a) The sample space is $\Omega = \{1, \ldots, 6\}$. The probability distribution is $p : \Omega \to [0, 1]$, where $p(3) = 2/7$ and $p(1) = p(2) = p(4) = p(5) = p(6) = 1/7$. Then $p(3) = 2 \cdot p(i)$ for each $i \in \Omega \setminus \{3\}$. Thus, the probability that the die shows 3 pips is twice as high as for every other number. To confirm that this is a probability distribution we need to check that the sum of probabilities of elementary events is 1:

$$\sum_{w \in \Omega} p(w) = p(1) + p(2) + \cdots + p(6) = 5 \cdot 1/7 + 2/7 = 1.$$

(b) Similarly, as in (a), the probability distribution for the blue die is $q : \Omega \to [0, 1]$, where $q(4) = 2/7$ and $q(1) = q(2) = q(3) = q(5) = q(6) = 1/7$. (For a full solution, you need to check that it has the required properties in the same way as we did for part (a).)

(c) The sample space for both dice is Ω^2, and each elementary event (x, y) occurs with probability $\text{Prob}(x, y) = p(x) \cdot q(y)$, because the outcomes of the blue and the red die are independent. Thus, we have a probability distribution $\text{Prob} : \Omega^2 \to [0, 1]$, where $\text{Prob}(x, y) = p(x) \cdot q(y)$. The event that the sum of pips is 7 is

$$E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

We have $\text{Prob}(3, 4) = p(3) \cdot q(4) = 2/7 \cdot 2/7 = 4/49$. For each $x \neq 3$ and $y \neq 4$ we have $\text{Prob}(x, y) = p(x) \cdot q(y) = 1/7 \cdot 1/7 = 1/49$. Hence,

$$\begin{aligned}\text{Prob}(E) &= \underbrace{p(1,6)}_{1/49} + \underbrace{p(2,5)}_{1/49} + \underbrace{p(3,4)}_{4/49} + \underbrace{p(4,3)}_{1/49} + \underbrace{p(5,2)}_{1/49} + \underbrace{p(6,1)}_{1/49} \\ &= 5 \cdot \frac{1}{49} + 1 \cdot \frac{4}{49} = \frac{9}{49}.\end{aligned}$$

Exercise 1.14 (a) Suppose we choose a vehicle at random among the 400 vehicles on the parking lot. Let R denote the event that it is red, and T that it is a truck. Then we have the following probabilities:

- $p(R) = 120/400$ (120 vehicles are red).
- $p(T) = 60/400$ (60 vehicles are trucks).
- $p(R \cap T) = 20/400$ (20 vehicles are red trucks).

Thus, the conditional probability that a vehicle is a truck, given that it is red is

$$p(T \mid R) = \frac{p(T \cap R)}{p(R)} = \frac{20/400}{120/400} = \frac{20}{120} = \frac{1}{6}.$$

Exercise 1.15

(a) Let $S \subseteq \{1, \ldots, n\}$ and B_S the event that exactly the bits in S are 1. If $|S| = 3$, then

$$p(B_S) = .51^3 \cdot .49^{n-3}.$$

Note that if S and S' are distinct sets of size 3, then B_S and $B_{S'}$ are disjoint events (i.e., they cannot both occur). Hence, the probability that there are exactly three 1s is

$$p\left(\bigcup_{\substack{S \subseteq \{1,\ldots,n\} \\ |S|=3}} p(B_S)\right) = \sum_{\substack{S \subseteq \{1,\ldots,n\} \\ |S|=3}} p(B_S) = \sum_{\substack{S \subseteq \{1,\ldots,n\} \\ |S|=3}} .51^3 \cdot .49^{n-3} = \binom{n}{3} .51^3 \cdot .49^{n-3}.$$

(b) $p(\text{at least one bit is 1}) = 1 - p(\text{all bits 0}) = 1 - .49^n$

(c) $p(\text{all bits same}) = p(\text{all bits 0 or all bits 1}) = p(\text{all bits 0}) + p(\text{all bits 1}) = .49^n + .51^n$.

Exercise 1.19 This answer is simple, if use the right "algorithm" to choose the integers $s_1, \ldots, s_n$:

1. First, we choose a set $\{x_1, \ldots, x_i\}$ of i distinct integers from S without ordering them.
2. Next, we order the integers in that set randomly, and let s_j be the j-th integer in the order. Thus, randomly ordering the elements from the set $\{x_1, \ldots, x_i\}$ yields the sequence $(s_1, \ldots, s_i)$.
3. Finally, we choose distinct integers $s_{i+1}, \ldots, s_n$ at random from the remaining elements in S.

Obviously, the above algorithm yields the desired distribution for $s_1, \ldots, s_n$, i.e., these are distinct random numbers, each chosen uniformly from S (without replacement).

Moreover, whether s_i is larger than $s_1, \ldots, s_{i-1}$ or not, is uniquely determined by the second step of the algorithm. Specifically, s_i is larger than $s_1, \ldots, s_{i-1}$ if and only if the largest element from $\{x_1, \ldots, x_i\}$ is put last in the order. This happens with probability $1/i$.

Hence, the probability that s_i is larger than $s_1, \ldots, s_{i-1}$ is exactly $1/i$.

2. Random Variables

2.1. Random Variables and Expectation

2.1.1. Definition of Random Variables

See Section 7.2.7 in the textbook.

Definition 2.1. *Let Ω be a sample space and $p : \Omega \to [0, 1]$ a probability distribution. A* random variable *over Ω is a function $X : \Omega \to V$, where $V \subseteq \mathbb{R}$.*

Example 2.2. Suppose we choose a random bit string of length 3 and are interested in the number of 1s it contains. The sample space is $\Omega = \{0, 1\}^3$, and the number of 1s in a random bit string is described by the random variable $X : \Omega \to \{0, 1, 2, 3\}$, where

- $X(000) = 0$,
- $X(001) = X(010) = X(100) = 1$,
- $X(110) = X(101) = X(011) = 2$, and
- $X(111) = 3$. ◀

See Section 7.2.7 in the textbook for additional examples.

2.1.2. Notation

Let r be any real number. We will write "$X = r$" to denote the event $E = \{w \in \Omega \mid X(w) = r\}$. I.e., E is the event that the random experiment yields an outcome that is assigned the value r by the random variable.

Consider the random variable X from Example 2.2, which counts the number of 1s in a bit string of length 3. Then the event "$X = 1$" is $\{001, 010, 100\}$.

This allows us to write $p(X = 1)$ for the probability that the bit string has length 1. Similar notation is used for other comparison operations, such as "$\leq$" or "$>$".

For example, by counting the elementary events $w \in \{0,1\}^3$ for which $X(w) \leq 1$, we obtain that the probability that a random bit string of length 3 has at most one 1 is

$$p(X \leq 1) = p(\{000, 001, 010, 100\}) = 4/8 = 1/2.$$

Formally, a random variable is a function that assigns each elementary event a real number. Therefore, a random variable is neither random nor a variable. Thus, at a first glance the term "random variable" seems misleading. But we can (and will) indeed think of random variables as random numbers. Notation of the form "$p(X = 5)$" facilitates this way of thinking.

We can make this formal, too: Recall that X is a function that maps elementary events in Ω to a set V of numbers. Then we can consider V as a sample space, and we can define a new probability distribution $p' : V \to [0,1]$ by letting $p'(r) = p(X = r)$. Now p' is really the probability distribution of a "random number" in V.

2.1.3. Expectation

See Section 7.4.2 in the textbook.

Definition 2.3. *The* expectation *of a random variable X is*[1]

$$E[X] = \sum_{w \in \Omega} p(w) \cdot X(w).$$

The expectation of a random variable is also called *mean* or *expected value*. If $E[X] = r$ then we also say that the value of X is r *in expectation*.

Example 2.4. Suppose we choose a bit string of length 2 at random. What is the expected number of 1s?

We define $\Omega = \{0,1\}^2$ and $X : \Omega \to \{0,1,2\}$, where $X(00) = 0$, $X(01) = X(10) = 1$, and $X(11) = 2$. Then

$$\begin{aligned} E[X] &= p(\{00\})X(00) + p(\{01\}) \cdot X(01) + p(\{10\}) \cdot X(10) + p(\{11\})X(11) \\ &= \frac{1}{4} \cdot (0 + 1 + 1 + 2) = 1. \end{aligned}$$

◀

Using Definition 2.3 to compute the expectation of a random variable is often tedious. Instead, we can use the following alternative characterization:

[1] As throughout the entire course, we assume here that Ω is countable. Otherwise, we wouldn't be able to sum over all elements in Ω and would need to use some form of integration.

Theorem 2.5. *Let Ω be a sample space and $X : \Omega \to V$ a random variable for a countable set $V \subseteq \mathbb{R}$. Then*

$$E[X] = \sum_{i \in V} i \cdot p(X = i).$$

Example 2.6. Suppose we roll two dice. What is the expected sum of pips?

Let $\Omega = \{1, \ldots, 6\}^2$, p the uniform distribution over Ω, and X the random variable that counts the sum of pips on the two dice. Formally, $X : \Omega \to \{2, \ldots, 12\}$, where $X(v_1, v_2) = v_1 + v_2$. We have

$$\begin{aligned} p(X = 2) &= p(\{11\}) = 1/36 & \qquad p(X = 12) &= p(\{66\}) = 1/36 \\ p(X = 3) &= p(\{12, 21\}) = 2/36 & \qquad p(X = 11) &= p(\{56, 65\}) = 2/36 \\ p(X = 4) &= p(\{13, 22, 31\}) = 3/36 & \qquad p(X = 10) &= p(\{46, 55, 65\}) = 3/36. \end{aligned}$$

Similarly, we obtain

$$\begin{aligned} p(X = 5) &= 4/36 & \qquad p(X = 9) &= 4/36 \\ p(X = 6) &= 5/36 & \qquad p(X = 8) &= 5/36 \\ p(X = 7) &= 6/36. \end{aligned}$$

Hence,

$$E[X] = \sum_{i=2}^{12} i \cdot p(X = i) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7. \quad \blacktriangleleft$$

2.1.4. Linearity of Expectation

As Example 2.6 showed, the task of computing the expected value of the sum of pips of two dice is very tedious. The following theorem, gives us a way of simplifying this task significantly. It is one of the most useful facts for the analysis of algorithms.

Theorem 2.7 (Linearity of Expectation)**.**

(a) Let $X_1, \ldots, X_n$ be random variables. Then $E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$.

(b) Let X be a random variable, and $a, b \in \mathbb{R}$. Then $E[a \cdot X + b] = a \cdot E[X] + b$.

Example 2.8. As in Example 2.6, suppose we roll two dice, and we want to compute the expected sum of pips. We now define two random variables, X_1 and X_2, where X_i denotes the result of the i-th die. Then the sum of pips is $X_1 + X_2$. By linearity of expectation $E[X_1 + X_2] = E[X_1] + E[X_2]$.

Obviously, $p(X_1 = v) = 1/6$ for each possible value $v \in \{1, \ldots, 6\}$. Thus,

$$E[X_1] = \sum_{i=1}^{6} i \cdot p(X_1 = i) = \sum_{i=1}^{6} i \cdot \frac{1}{6}.$$ ◀

2.1.5. Indicator Random Variables

A common technique to determine the expectation of a random variable is to split it into a sum $X_1 + \cdots + X_n$ of multiple random variables, where each X_i, $i \in \{1, \ldots, n\}$, is either 0 or 1. Such random variables are called *indicator random variables*, or *0/1-random variables*.

Note that an indicator random variable X has expectation

$$E[X] = 0 \cdot p(X = 0) + 1 \cdot p(X = 1) = p(X = 1).$$

Fact 2.9. *An indicator random variable X has expectation $E[X] = p(X = 1)$.*

Example 2.10. Suppose we roll a die n times. What is the expected number of times we obtain a 6?

Let X_i be the random variable that takes value 1 if the i-th die roll shows a 6, and 0 if it shows a number in $\{1, \ldots, 5\}$. Then

$$E[X_i] = p(X_i = 1) = \frac{1}{6}.$$

Moreover, $X = X_1 + \cdots + X_n$ is exactly the number of 6's we see. By linearity of expectation we have

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{6} = \frac{n}{6}.$$

Hence, the expected number of 6's observed in n die rolls is $n/6$. ◀

Example 2.11. At a wedding party, each of the n guests is assigned a seat. Then everybody enters the dance floor, and after the dance, returns to a randomly chosen seat. How many guests return to their assigned seats (in expectation)?

Suppose the seats and guests are numbered $1, \ldots, n$, and guest i is assigned seat i. For each $i \in \{1, \ldots, n\}$, let $X_i = 1$ if guest i returns from the dance to their assigned seat, and otherwise $X_i = 0$. Then

$$E[X_i] = p(X_i = 1) = 1/n.$$

Now let X be the number of guests that return to their assigned seats. Then

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n \cdot \frac{1}{n} = 1.$$

Thus, in expectation exactly one guest returns to their assigned seat. ◀

Example 2.12. Consider the following algorithm to find the largest element in an array $A[1..n]$.

```
   Input: An integer n ≥ 0 and an array A[1..n] of positive integers.
1  max := −∞
2  for j := 1 . . . n do
3  |  if A[j] > max then
4  |  |  max := A[j]
5  |  end
6  end
7  print(max)
```

Assume that the array stores n distinct positive integers in random order. We are interested in the expected number of times the algorithm executes line 4.

Let X be the *random variable* that denotes the number of times the algorithm executes line 4. For each $i \in \{1, \ldots, n\}$, let X_i be the indicator random variable that has value 1 if and only if line 4 is executed in the i-th iteration of the loop, i.e., when $j = i$. Otherwise, $X_i = 0$. Then $X = X_1 + \cdots + X_n$. By linearity of expectation

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = \sum_{i=1}^{n} p(X_i = 1). \tag{2.1}$$

Hence, we need to determine the probability of the event "$X_i = 1$". This even occurs, if and only if $A[i]$ is larger than all of $A[1], \ldots, A[i-1]$. To determine the probability of this event, we can first fix i arbitrary distinct integers, and then randomly distribute them over $A[1], \ldots, A[i]$ (see also Section 1.3.5). No matter what i integers we choose, the probability that $A[i]$ is assigned the largest one among them is exactly $1/i$. Hence,

$$p(X_i = 1) = \frac{1}{i},$$

and thus by eq. (2.1),

$$E[X] = \sum_{i=1}^{n} \frac{1}{i}$$

This sum is called *harmonic series*, and the following inequalities can be used to bound it:

$$\ln(n+1) \leq \sum_{i=1}^{n} \frac{1}{i} \leq (\ln n) + 1, \tag{2.2}$$

where $\ln n$ is the natural logarithm (i.e., base e logarithm) of n. Therefore, we have

$$\ln(n+1) \leq E[X] \leq (\ln n) + 1. \qquad \blacktriangleleft$$

2.1.6. The Geometric Distribution

See Section 7.4.5. in the textbook.

Many of the random variables we analyze are running times of randomized algorithms. Such random variables are usually positive integers. If the random variable takes only non-negative integer values, then there is a very useful characterization of its expectation.

Theorem 2.13. *Let $X : \Omega \to \mathbb{N}_0$ be a random variable. If $E[X]$ exists, then*

$$E[X] = \sum_{i=0}^{\infty} p(X > i).$$

Proof.

$$\begin{aligned}
E[X] &= \sum_{k=0}^{\infty} k \cdot p(X = k) = \sum_{k=1}^{\infty} k \cdot p(X = k) \\
&= p(X = 1) + 2 \cdot p(X = 2) + 3 \cdot p(X = 3) + \ldots \\
&= \sum_{k=1}^{\infty} p(X = k) + \sum_{k=2}^{\infty} p(X = k) + \sum_{k=3}^{\infty} p(X = k) + \ldots \\
&= \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} p(X = k) = \sum_{i=1}^{\infty} p(X \geq i) \\
&= \sum_{i=1}^{\infty} p(X > i - 1) = \sum_{i=0}^{\infty} p(X > i).
\end{aligned}$$

□

Note that Theorem 2.13 is only true if X takes only non-negative integer values. The theorem is useful in the analysis of the so-called *geometric distribution*.

Suppose we keep rolling a die until it shows a 6. How often do we have to roll it in expectation? Many people will intuitively give the correct answer: 6 times.

Let us do the math, but let us be more general. Imagine, we perform a random experiment whose outcome is either F or S (as in "fail" and "success"). Let q be the probability that the outcome of the experiment is S. (In the die rolling example, S corresponds to rolling a 6, and so $q = p(S) = 1/6$.) We now repeat the random experiment, until we obtain the first S (e.g., until the die shows a 6 for the first time). Let X be the number of times we have to perform the random experiment. What is $E[X]$?

To apply Theorem 2.13, we will now compute the probability of the event "$X > k$" for each integer $k \geq 0$. Let S_i be the event that the i-th outcome is S. Note that $X > k$ is true if and only if the first i-th outcomes are all F, i.e., S_i does not occur for any $i \in \{1, \ldots, k\}$. Hence,

$$p(X > k) = p\big(\overline{S_1} \cap \overline{S_2} \cap \cdots \cap \overline{S_k}\big) = p\big(\overline{S_1}\big) \cdot p\big(\overline{S_2}\big) \ldots p\big(\overline{S_k}\big)$$
$$= \underbrace{(1-q) \cdot (1-q) \cdots (1-q)}_{k \text{ times}} = (1-q)^k.$$

Thus, from Theorem 2.13 we obtain

$$E[X] = \sum_{k=0}^{\infty} p(X > k) = \sum_{k=0}^{\infty} (1-q)^k = \frac{1}{q}.$$

(The last identity follows from the *geometric series*, which is for any r

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}, \tag{2.3}$$

and which, for $0 < r < 1$, converges to $1/(1-r)$ if we let n go to ∞.)

Note that above we computed the probability of the event $X > k$. Similarly, it is not hard to see that $p(X = k) = (1-q)^{k-1} \cdot q$, because the event "$X = k$" occurs exactly if the first $k-1$ outcomes are all F, and the k-th outcome is S:

$$p(X = k) = p\big(\overline{S_1} \cap \cdots \cap \overline{S_{k-1}} \cap S_k\big) = p\big(\overline{S_1}\big) \cdots p\big(\overline{S_{k-1}}\big) \cdot p(S_k) = (1-q)^{k-1} \cdot q.$$

Definition 2.14 (Geometric Distribution)**.** *A random variable $X : \Omega \to \mathbb{N}$ is geometrically distributed with parameter q, where $0 \leq q \leq 1$, if $p(X = k) = (1-q)^{k-1} \cdot q$ for each $k \in \mathbb{N}$.*

Above we proved the following result:

Theorem 2.15. *If X is geometrically distributed with parameter $q > 0$, then $p(X > k) = (1-q)^k$ for any $k \in \mathbb{N}$, and $E[X] = 1/q$.*

Example 2.16. Consider an array of size n such that each array entry is either 0 or 1. Suppose we repeatedly query random array entries until we find the first 1. (This corresponds to finding an entry in a database with a specific property.) Suppose exactly k of the n array entries are 1. Then the probability of finding a 1 in one specific query is $q = k/n$. The number of queries we need until we query a 1 (i.e., until we have success) is a geometrically distributed random variable with parameter q. Hence, the expected number of queries is $1/q = n/k$. ◀

2.1.7. Repeat Until Multiple Successes

Suppose we repeatedly roll a die, until we have rolled a 6 k times. How many die rolls do we perform in expectation?

For each $i \in \{1, \dots, k\}$, let X_i be the number of die rolls until we roll the i-th 6, starting from the point when we have rolled $i-1$ 6s. Then $X = X_1 + \cdots + X_k$ is the total number of die rolls until we have rolled k 6s. Since each X_i is geometrically distributed with parameter $1/6$, we have $E[X_i] = 6$, and thus

$$E[X] = E[X_1 + \cdots + X_k] = E[X_1] + \cdots + E[X_k] = 6k.$$

2.1.8. Repeat Until Success or Bound

Now assume we repeatedly flip a coin until either we have flipped heads, or until we have performed k flips (whichever happens first). Let X be the total number of flips we perform. Then according to Theorem 2.13,

$$E[X] = \sum_{i=0}^{\infty} p(X > i).$$

Since we always stop after k flips, we have $p(X > k) = 0$. Similarly, for each $i \geq k$, we have $p(X > i) = 0$. On the other hand, if $i < k$, then the event $X > i$ occurs if and only if the first i flips are all tails. Hence, for $i < k$ we have $p(X > i) = (1/2)^i$. Therefore, we obtain

$$E[X] = \sum_{i=0}^{\infty} p(X > i) = \sum_{i=0}^{k-1} p(X > i) = \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^i \overset{(2.3)}{=} \frac{1 - (1/2)^k}{1 - 1/2} = 2 - \frac{1}{2^{k-1}}.$$

2.1.9. Coupon Collecting

Suppose each cereal box contains a coupon of a certain type, which is randomly chosen among n types. Our goal is to collect at least one coupon of each type. How many cereal boxes do we have to buy in expectation?

This is the same as the expected number of balls we need to throw at random into n bins, until each bin contains at least one ball (we say a bin is *covered* if it contains a ball). Let X be the number of balls needed, until all n bins are covered. Consider some point in time when exactly $i-1$ bins are covered. Let X_i be the number of additional balls we throw into random bins until i bins are covered. Then $X = X_1 + \cdots + X_n$.

The probability that a ball is thrown into an empty bin, when $i-1$ bins are covered, is $(n-i+1)/n$. Thus, X_i is geometrically distributed with parameter $q=(n-i+1)/n$ and $E[X_i]=n/(n-i+1)$. We get

$$E[X]=\sum_{i=1}^{n}E[X_i]=\sum_{i=1}^{n}\frac{n}{n-i+1}=n\cdot\sum_{i=1}^{n}\frac{1}{i}.$$

Recall that the sum $1+1/2+\cdots+1/n$ is the harmonic series, and can be bounded from below by $\ln(n+1)$ and from above by $(\ln n)+1$—see eq. (2.2).

Thus, we have

$$n\ln n\leq E[X]\leq n\ln n+n.$$

2.1.10. Independent and Dependent Random Variables

Recall that two events A and B are independent, if $p(A\cap B)=p(A)\cdot p(B)$. Similarly, two *random variables* are independent, if all events associated with each of these two random variables do not depend on each other. Precisely, random variables X and Y are independent, if for all values x and y,

$$p(X=x\wedge Y=y) \;=\; p(X=x)\cdot p(Y=y).$$

Observe that if the above identity is true, then for all x and y:

$$p(X=x)=p(X=x\mid Y=y)\quad\text{and}\quad p(Y=y)=p(Y=y\mid X=x).$$

For example, if we roll two dice, a red one and a blue one. Let X denote the result of the red die, and Y the result of the blue die. Then for each value $x=\in\{1,\ldots,6\}$, $p(X=x)=1/6$, and for each value $y\in\{1,\ldots,6\}$, $p(Y=y)=1/6$. Moreover, $p(X=x\wedge Y=y)=1/36=p(X=x)\cdot p(Y=y)$. Hence, X and Y are independent.

Now consider the following random experiment: First, we roll a red die, and let R denote the result. Now we roll the blue die. If the blue die also shows R pips (i.e., we roll the same number as the red die shows), then we roll the blue die again. We repeat rolling the blue die, until it shows a number that is different from R. Let B be the number of pips the blue die shows in the end.

Then R and B are not independent: Obviously, we have $p(R=r)=1/6$ for every value $r\in\{1,\ldots,6\}$, because R is just the result of a single die roll. We also have $p(B=b)=1/6$ for each $b\in\{1,\ldots,6\}$. To see that this is true, observe that each outcome for B is equally likely (as long as we have not fixed R). Hence, B must be uniformly distributed over $\{1,\ldots,6\}$, and thus each result occurs with probability $1/6$. But, since $R=1$ and $B=1$ cannot both occur, we have $p(R=1\wedge B=1)=0$, which is not equal to $p(R=1)\cdot p(B=1)=1/36$.

It is important to note that linearity of expectation is still true for dependent random variables. For the two random variables R and B, we still have

$$E[R + B] = E[R] + E[B] = 3.5 + 3.5 = 7,$$

even though R and B are dependent.

But only if two random variables are independent, then something similar holds for the product of two random variables. In that case, the expectation of the product of two independent random variables is equal to the product of their expectations:

Theorem 2.17. *If X and Y are two independent random variables and their expectations exist, then*

$$E[X \cdot Y] = E[X] \cdot E[Y].$$

Note that the above identity is in general not true, if X and Y are dependent. The above theorem can be generalized to more than two mutually independent random variables.

Theorem 2.18. *If $X_1, \ldots, X_k$ are mutually independent random variables and their expectations exist, then*

$$E[X_1 \cdot \ldots \cdot X_k] = E[X_1] \cdot \ldots \cdot E[X_k].$$

2.2. Tail Bounds

A *tail bound* is a bound of the form "$p(X > \ldots) < \ldots$", i.e., it bounds the probability that a random variable is very large (typically much larger than its expectation).

2.2.1. Tail Bounds for the Geometric Distribution

Sometimes we can determine the probability of $p(X > \ldots)$ exactly. For example, let X denote the number of die rolls until the first 6 is rolled. Let A_j denote the event that the j-th die roll is a 6. Then

$$p(X > i) = p\big(\overline{A_1} \wedge \overline{A_2} \wedge \cdots \wedge \overline{A_i}\big) = \left(\frac{5}{6}\right)^i.$$

More generally, let X be a geometrically distributed random variable with parameter q. Recall that X counts the total number or repetitions of a random experiment with success probability q, until success occurs for the first time. Hence, "$X > i$" is the event that the first i outcomes are not successful, and thus has probability

$$p(X > i) = (1 - q)^i.$$

To bound this probability from above for large value of i, we can apply the following very useful inequality:

$$\left(1 - \frac{1}{x}\right)^x < \frac{1}{e} \text{ for any } x \geq 1. \tag{2.4}$$

Now recall that $E[X] = 1/q$, because X is geometrically distributed. Thus, for any constant $c > 0$, we can bound the probability that X is a factor c larger than its expectation as follows.

$$\begin{aligned} p(X > \lceil c \cdot E[X] \rceil) &= p(X > \lceil c \cdot 1/q \rceil) = (1-q)^{\lceil c \cdot (1/q) \rceil} \leq (1-q)^{c \cdot (1/q)} \\ &= \left(\left(1 - \frac{1}{1/q}\right)^{1/q}\right)^c < \left(\frac{1}{e}\right)^c. \end{aligned} \tag{2.5}$$

2.2.2. Coupon Collecting Revisited

Consider again the problem from Section 2.1.9: There are n bins, and we keep throwing balls into random bins, until all bins are covered. Let X be the number of balls needed. In Section 2.1.9 we have seen that $E[X] \leq n \ln n + n$. Now we would like to derive a tail bound for X, i.e., upper bound the probability that X is much larger than $E[X]$. For example, let us bound $p(X > 2n \ln n)$.

Random variable X is not geometrically distributed, so we cannot use a similar calculation as the one in eq. (2.5) directly. Instead, we consider each bin i individually: Let X_i denote the number of balls thrown into the bins until bin i has received at least one ball. Whenever we throw a ball into one of the n bins, the ("success") probability that bin i receives the ball is $1/n$. Hence, X_i is geometrically distributed with parameter $1/n$, and thus:

$$p(X_i > \lceil 2n \ln n \rceil) \leq (1-1/n)^{2n \ln n} = \left((1-1/n)^n\right)^{2 \ln n} < \left(\frac{1}{e}\right)^{2 \ln n} = \left(\left(\frac{1}{e}\right)^{\ln n}\right)^2 = \frac{1}{n^2}.$$

For the ease of description, assume now that $2n \ln 2$ is an integer (so $\lceil 2n \ln n \rceil = 2n \ln n$). Now recall that we want to determine an upper bound for the probability of event "$X > 2n \ln n$". This event occurs, if at least one bin is not covered by the first $2n \ln n$ balls thrown into the bins. I.e., there is a bin i such that $X_i > 2n \ln n$. Hence, the event "$X > 2n \ln n$" is equivalent to the event "$X_1 > 2n \ln n \vee X_2 > 2n \ln n \vee \cdots \vee X_n > 2n \ln n$". We will now apply the union bound:

$$\begin{aligned} p(X > 2n \ln n) &= p(X_1 > 2n \ln n \vee \cdots \vee X_n > 2n \ln n) \\ &\leq p(X_1 > 2n \ln n) + \cdots + p(X_n > 2n \ln n) \\ &< \underbrace{\frac{1}{n^2} + \cdots + \frac{1}{n^2}}_{n \text{ times}} = n \cdot \frac{1}{n^2} = \frac{1}{n}. \end{aligned}$$

2.2.3. Markov's Inequality

The tail bounds we derived above are specific to the geometric distribution. Often we do not have enough information about the distribution of a random variable X, to derive such strong bounds. The following tail bound, called Markov's Inequality, is much weaker, but applies to all random variables.

Theorem 2.19 (Markov's Inequality)**.** *Let V be a set of non-negative real numbers, and let $X : \Omega \to V$ be a random variable. Then for all $t \in \mathbb{R}_{>0}$,*

$$p(X \geq t) \leq \frac{E[X]}{t}.$$

Note that Markov's Inequality only holds for *non-negative* random variables.

Example 2.20. Suppose that in a CPSC 251 class of 150 students, the average score on the final exam was 75%. What is an upper bound on the number of students, who scored at least 90%?

To answer this question, we choose one of the 150 students at random, and let X denote the student's score on the final exam. Then we know that $E[X] = 0.75$. Since all scores are positive, X is a non-negative random variable. Hence, by Markov's Inequality:

$$p(X \geq 0.9) \leq E[X]/0.9 = 0.75/0.9 = 0.8333\ldots.$$

Thus, at most 83% of the 150 students scored at least 90%. This makes at most 125 students. ◀

Example 2.21. Consider the algorithm from Example 2.12. As we showed in this example, the algorithm executes line 4 at most $(\ln n) + 1$ times. But in the worst-case, it executes that line n times. What is the probability that the worst-case occurs?

Let X be the total number of times that line gets executed. Then X is a non-negative random variable, and $E[X] \leq (\ln n) + 1$. Hence, by Markov's Inequality

$$p(X \geq n) \leq \frac{E[X]}{n} = \frac{(\ln n) + 1}{n}.$$ ◀

We note that Markov's Inequality often leads to *weak* tail bounds. For example, suppose we repeatedly roll a die until we roll a 6. The probability that a single die roll yields a 6 is 1/6. If X is the number of die rolls until this happens is geometrically distributed with parameter 1/6, and the probability that we need more than n die rolls is

$$p(X > n) = (5/6)^n.$$

For $n = 20$ this is less than 0.026, and for $n = 50$ it is less than 0.00011.

We could use Markov's Inequality instead. Random variable X is non-negative and $E[X] = 6$. Hence,

$$p(X > n) < \frac{E[X]}{n} = \frac{6}{n}.$$

Now, for $n = 20$ we obtain an upper bound of $6/20 = 0.3$ and for $n = 50$ a bound of 0.12. Overall, for growing n, the upper bound using Markov's Inequality shrinks much slower than using the exact calculation above.

2.2.4. Variance, Standard Deviation, and Chebyshev's Inequality

Note

This section is not part of the curriculum, and will not be tested.

Variance, standard deviation, and Chebyshev's inequality are discussed in the textbook, Sections 7.4.7 and 7.4.8. Here, we only summarize the main definitions and results. (Theorem 2.24 below is an immediate generalization of Theorem 7 on p. 515 in the textbook.) For examples, see the relevant textbook sections.

Definition 2.22. *Let $X : \Omega \to V$ be a random variable. The* variance *of X is*

$$\mathrm{Var}\,[X] = E\Big[\big(X - E[X]\big)^2\Big].$$

The standard deviation *of X is*

$$\sigma(X) = \sqrt{\mathrm{Var}\,[X]}.$$

Observe that $E[X]$ is a fixed number (and not a random variable), while $Z = (X - E[X])^2$ is a random variable. The variance of X is simply the expectation of Z. Therefore, we obtain the

following identities:

$$\begin{aligned}
\operatorname{Var}[X] &= E\Big[(X - E[X])^2\Big] \\
&= \sum_{w\in\Omega} \big(X(w) - E[X]\big)^2 p(w) \\
&= \sum_{w\in\Omega} \big(X(w)^2 - 2X(w)E[X] + E[X]^2\big) \cdot p(w) \\
&= \underbrace{\sum_{w\in\Omega} X(w)^2 p(w)}_{=E[X^2]} - \sum_{w\in\Omega} 2X(w)E[X]p(w) + E[X]^2 \\
&= E[X^2] - 2E[X] \cdot \underbrace{\sum_{w\in\Omega} X(w)p(w)}_{=E[X]} + E[X]^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2.
\end{aligned}$$

This yields the following:

Theorem 2.23. *For any random variable X,*

$$\operatorname{Var}[X] = E[X^2] - E[X]^2.$$

Random variables $X_1, \ldots, X_k$ are *pairwise independent*, if any two of them are independent.

Theorem 2.24. *Let $X_1, \ldots, X_k$ be pairwise independent random variables. Then*

$$\operatorname{Var}[X_1 + \cdots + X_n] = \operatorname{Var}[X_1] + \cdots + \operatorname{Var}[X_n].$$

Theorem 2.25 (Chebyshev's Inequality)**.** *For any random variable X and any $t > 0$,*

$$p(|X - E[X]| \geq t) \leq \frac{\operatorname{Var}[X]}{t^2}.$$

2.3. Algorithmic Applications

2.3.1. Average Case Analysis

Often, we analyze the running time of algorithms by considering only the worst-case input. That is, we assume an input that causes an algorithm to run as long as possible. But in many scenarios, that worst-case very rarely occurs.

An alternative approach is to perform an *average case analysis*, where we assume a random input, and we compute the expected running time of an algorithm.

Consider the following *linear search* algorithm. It takes as input a list $L[1] \dots L[n]$ of n values, as well as a value v. The algorithm then searches through all list elements, in order to determine if v is in the list. Once it finds the element v in an entry $L[i]$, it returns the index i of that entry. If v is not in the list, the algorithm outputs "not found".

```
Input: A list L[1], ..., L[n] and a value v.
for i = 1 ... n do
    if L[i] = v then
        print(i)
        return
    end
end
print("not found")
```

For the purpose of this analysis, we make the following assumptions:

1. v appears in the list with some probability q, where $0 \leq q \leq 1$;
2. v appears at most once in the list; and
3. if v is in the list, then its position is uniformly distributed over all n list elements.

We are interested in the expected number of iterations of the for-loop.

Let random variable X denote the number of for-loop iterations. Moreover, let A be the event that "v appears in the list", and for $i \in \{1, \dots, n\}$, we let B_i be the event that $L[i] = v$.

We have the following probabilities for each $i \in \{1, \dots, n\}$.

$$\begin{aligned} p(A) &= q && (v \text{ is in the list with probability } q) \\ p(B_i \mid A) &= \frac{1}{n} && (\text{if } v \text{ is in the list, it is in each position with probability } 1/n) \\ p\big(B_i \mid \overline{A}\big) &= 0 && (\text{if } v \text{ is not in the list, it is not in position } i). \end{aligned}$$

Using the total probability theorem, we obtain

$$p(B_i) = p(B_i \mid A) \cdot p(A) + p\big(B_i \mid \overline{A}\big) \cdot p\big(\overline{A}\big) = \frac{1}{n} \cdot q + 0 \cdot (1 - q) = \frac{q}{n}.$$

Observe that if B_i occurs, then the for-loop ends after i iterations, so $X = i$. Moreover, if $\overline{A}$ occurs, then the element is not in the list, and all n iterations of the for-loop are executed. So

$X = n$. Thus, for $i \in \{1, \ldots, n-1\}$, event B_i is equivalent to event $X = i$, and event B_n is equivalent to $(X = i \vee \overline{A})$. We obtain

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} i \cdot p(X = i) = \sum_{i=1}^{n-1} i \cdot p(X = i) + n \cdot p(X = n) = \sum_{i=1}^{n-1} i \cdot p(B_i) + n \cdot p\big(B_n \vee \overline{A}\big) \\ &= \sum_{i=1}^{n-1} i \cdot p(B_i) + n \cdot p(B_n) + n \cdot p\big(\overline{A}\big) = \sum_{i=1}^{n} i \cdot p(B_i) + n(1-q) = \sum_{i=1}^{n} i \cdot \frac{q}{n} + n(1-q) \\ &= \frac{q}{n} \cdot \frac{n(n+1)}{2} + (1-q) \cdot n = q \cdot \frac{n+1}{2} + (1-q) \cdot n. \end{aligned}$$

(We used the well-known fact that $1 + \cdots + n = n(n+1)/2$.)

Observe that if $q = 0$, then v is in the list with probability 0. For that case, the above result indicates that $E[X] = n$. This makes sense, because we need to search the entire list. On the other hand, if $q = 1$, v is in the list with probability 1. For that case we obtain $E[X] = (n+1)/2$. This also makes sense, because we expect v to be in the middle of the list.

2.3.2. Monte Carlo Algorithms

See Section 7.2.9 in the textbook.

2.3.3. Las Vegas Algorithms

Contrary to a Monte Carlo algorithm, a Las Vegas algorithm may not produce an erroneous answer. However, now the running time is a random variable. Usually we are interested in the expected running time, or in a tail bound for the running time.

Assume that our programming language provides a method `random_bit()`, which returns a uniformly distributed random bit. Multiple calls of the method return independent random bits. But now we want to compute a random number in $\{0, 1, 2\}$.

We can use the following Las Vegas algorithm:

Observe that the number v computed in line 4 is uniformly distributed in $\{0, 1, 2, 3\}$. Thus, given that $v \neq 3$, v is uniformly distributed over $\{0, 1, 2\}$. I.e., for each $i \in \{0, 1, 2, 3\}$, we have $p(v = i \mid v \neq 3) = 1/3$. It follows that the algorithm returns each number $v \in \{0, 1, 2\}$ with equal probability $1/3$.

But the running time of the algorithm is a random variable. Let X denote the number of loop iterations. In each such iteration, with probability $1/4$ the algorithm calculates $v = 3$, and has to perform another iteration. The "success" probability of choosing $v \in \{0, 1, 2\}$ is $3/4$. Thus, X is geometrically distributed with parameter $3/4$, and so the number of loop iterations has expectation $E[X] = 4/3$.

```
repeat
    b_0 = random_bit()
    b_1 = random_bit()
    v = b_0 + 2b_1
    // This maps b_0 b_1 to an integer v ∈ {0,1,2,3} as follows:  00 → 0,
       01 → 1, 10 → 2, 11 → 3.
until v ≠ 3
return v
```

2.3.4. Converting Las Vegas to Monte Carlo

Usually we can convert a Las Vegas algorithm to a Monte Carlo one. Consider a Las Vegas algorithm A_{LV} (which always returns the correct answer) and let X denote the running time of that algorithm. We let A_{LV} run for $c \cdot E[X]$ steps, where $c > 1$ is some parameter. If A_{LV} does not terminate after $c \cdot E[X]$ steps, we simply return an arbitrary answer (which may be incorrect). By Markov Inequality, the probability that this happens is

$$p(X > c \cdot E[X]) < \frac{1}{c}.$$

Thus, we now have a Monte Carlo algorithm with worst-case running time $c \cdot E[X]$, and error probability at most $1/c$.

2.4. Exercises

2.1 In a lottery 1000 tickets are being sold for \$1, each. Four winning tickets are chosen uniformly at random. The prizes for the winning tickets are \$500, \$100, \$50, and \$10, respectively. If you purchase two tickets, what is your expected net gain?

expectation

2.2 Alice and Bob play a game in multiple rounds: In each round, each player flips a coin. If one player flips heads and the other tails, then heads wins the round. If both players flip tails or both flip heads, then Alice wins (Bob is being nice). Alice and Bob keep playing until one of them has won at least two rounds. How many rounds do they play in expectation?

expectation

2.3 Suppose we throw n balls into m bins, such that the ball locations are uniform and mutually independent. For (a) and (b) use the fact that the bin choices for different balls are mutually independent.

independence, random variables, expectation, indicator random variables

(a) What is the probability that the first 3 balls collide?

(b) What is the probability that all balls land in the same bin?

(c) What is the expected total number of balls in bins 1 and 2 combined?

2.4 Suppose you buy a \$5 Lotto Max ticket, which allows you to bet on three lines of 7 numbers out of 50. (Recall that the draw is 7 out of 50.) Suppose for 7 correct numbers on the same line you win \$50,000,000 and for 6 correct numbers on the same line you win \$100,000. For fewer than 6 correct numbers you don't win anything.

expectation, linearity of expectation

What is your expected surplus?

2.5 We assign each card the following point values: The number of points of a pip card (ace, 2,. . . ,10) is equal to the number of its pips (an ace is assigned 1 point), and each face card (jack, queen, king) is assigned 10 points.

linearity of expectation

Suppose we draw two cards from a deck of 52 at random.

(a) What is the expected point value of the first card?

(b) What is the expected sum of points of both cards?

linearity of expectation

2.6 Let A be an array that stores n distinct numbers. The efficiency of some algorithms depends on how well the array is sorted. One measure for "sortedness" is the number of *inversions*. An inversion is a pair $(i, j) \in \{0, \ldots, n-1\}^2$, where $i < j$ but $A[i] > A[j]$.

Suppose A stores a random permutation of $\{1, \ldots, n\}$. What is the expected number of inversions?

expectation, independence

2.7 Let b be a sequence of bits. A *run* is a maximal consecutive sub-sequence of 1s. For example, if $b = (1101011101111100)$, then b contains 4 runs, one of length 2, one of length 1, one of length 3, and one of length 5.

Let $q \in [0, 1]$. Suppose b contains n bits, and each bit is chosen independently at random, and it is 1 with probability q.

(a) What is the expected number of runs in b?

(b) Does the expectation change, if the bits are not mutually independent?

(c) Describe what events have to be independent so that your analysis in part (a) is valid.

expectation

2.8 Let A be an array of size n that has exactly one array entry with value 1, and all other array entries have value 0. Consider the following randomized algorithm to find the array entry with value 1:

(1) Choose an index i uniformly at random among all indices in $\{0, \ldots, n-1\}$ that were not chosen before. (2) Query $A[i]$. (3) If $A[i] = 1$, then output i, otherwise go to (1).

How many queries does the algorithm perform in expectation?

randomized algorithm, expectation, geometric distribution

2.9 Consider an array $A[0..n-1]$ of n bits, where n is even. Suppose exactly $n/2$ bits have value 1. Consider the following randomized algorithm that tries to find an array entry with value 1:

> *Choose an index $i \in \{0, \ldots, n-1\}$ at random. Query $A[i]$. If $A[i] = 1$, then print i and stop. Otherwise, start over.*

(a) What is the probability that the algorithm stops after the first query?

(b) What is the probability that the algorithm stops after exactly k queries for $k \geq 2$?

(c) What is the expected number of queries performed by this algorithm?

2.10 (a) Suppose we repeatedly roll three fair dice (all of them together) until the sum of all pips is even. How many times do we have to roll in expectation?

geometric distribution

(b) How does the answer change, if we also stop after rolling ten times, if no sum of even pips shows before that.

2.11 Let A be a randomized algorithm that takes as input a positive integer x and outputs either "prime" or "composite". Assume further, that if the algorithm outputs "composite", then x is indeed composite. But if it outputs "prime", then x may also be composite, but only with probability at most $1-1/n$, where n is the number of bits needed to describe x. Moreover, the running time of the algorithm is at most $T(n)$ for some function T.

geometric distribution, Monte Carlo algorithm, Las Vegas algorithm

(a) Describe a randomized algorithm that gives the wrong answer with probability at most $1/2$, no matter how big n is.

(b) Now assume that A always gives one of 3 answers, "prime", "composite", or "?". If its output is "prime" or "composite", then that is the correct answer. Suppose the probability that the algorithm outputs "?" is at most $1 - 1/n$. Give an algorithm that **always** gives the correct output (i.e., never outputs "?"), but whose running time is a random variable. What is the expectation of that random variable? (Recall that $(1-1/x)^x \leq 1/e$ for any $x \geq 1$.)

2.12 Suppose we fill an array $A[0], \ldots, A[n-1]$ with a random permutation. I.e., each array entry is assigned a random number in $\{0, \ldots n-1\}$, such that all array entries are distinct. What is the expected number of indices $i \in \{0, 1 \ldots, n-1\}$, such that $A[i] = i$?

linearity of expectation

2.13 Let n be some positive integer. A subsequence of a sequence $(a_1, \ldots, a_k)$ is a sequence that can be obtained by removing zero or more elements from $(a_1, \ldots, a_k)$. For example, $(3, 7, 5)$ is a subsequence of $(9, 8, 5, \underline{3}, 8, 3, 5, 9, 4, \underline{7}, 1, 4, \underline{5})$, but not of $(9, 8, 5, 3, 8, 3, 5, 9, 4, 7, 1, 4)$.

linearity of expectation, geometric distribution

Suppose we repeatedly generate random integers in $\{1, \ldots, n\}$, until we have obtained a sequence, of which $(1, 2, 3, \ldots, n)$ is a subsequence. What is the expected number of random integers produced?

2.14 Suppose Alice and Bob play a game, where in each round, each of them rolls a die. If both dice show the same result, then there is a tie.

bounded geometric distribution, multiple successes

Otherwise, the player with the higher die roll wins.

(a) Alice and Bob play the game, until one of them has won a round. Determine the total expected number of die rolls (not rounds).

(b) What is the expected number of die rolls, if they stop when Alice has won at least five rounds?

(c) *Challenge question:* What is the expected number of die rolls, if they stop as soon as one of them has won two rounds?

2.15 Suppose we repeatedly generate random numbers in $\{1, \ldots, n\}$, until we have generated each number in that set at least once. Let X denote the total number of random numbers generated.

Coupon collecting, tail bounds, Markov's Inequality

(a) What is $p(X = n)$?

(b) Use Markov's Inequality to determine an upper bound for $p(X > 2 \cdot n(\ln n + 1))$.

(c) Use a more sophisticated analysis to prove that $p(X > 2 \cdot n \ln n) < 1/n$.

2.16 In this question, we throw balls into n bins, and for each ball the bin is chosen uniformly and independently at random.

tail bounds, geometric distribution

(a) Give a value k, such that if we throw k balls into 100 bins, all bins receive at least one ball with probability **at least** $1 - 1/10^6$

(b) For any constant $c > 1$, determine an upper bound for the probability that not all n bins are covered, if we throw $c\ln n$ balls.

2.17 For any real number $t \geq 1$ find a non-negative random variable X, such that $p(X \geq t \cdot E[X]) = 1/t$.

Markov's Inequality

2.18 Show that Markov's Inequality is in general not true for random variables that may take negative values.

Markov's Inequality

2.19 Suppose X and Y are independent random variables. Prove the following statements:

Variance, expectation, independence

(a) $E[X \cdot Y] = E[X] \cdot E[Y]$, and

(b) $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

2.20 Give *dependent* random variables X and Y, for which the identities in Exercise 2.19 are not true.

Variance, expectation, independence

2.21 Suppose we roll a die n times, and let X be the sum of all die roll results.

Markov's Inequality, Chebyshev's Inequality

1. Use Markov's Inequality to prove an upper bound for $p(X > 3.5(n + \sqrt{n}))$.
2. Use Chebyshev's Inequality to prove an upper bound for the same probability.
3. Challenge question: Can you find a better upper bound than what you can prove with Chebyshev's Inequality?

Chebyshev's Inequality, probability amplification, randomness, Monte Carlo algorithms, pairwise independence

2.22 *Challenge question.*[2] Let p be a prime, and let $x, x' \in \{0, \ldots, p-1\}$ be two distinct integers.

It is well known that if we choose $a, b \in \{0, \ldots, p-1\}$ uniformly and independently at random, then $(ax + b) \bmod p$ and $(ax' + b) \bmod p$ are independent.

Now consider an array $A[0] \ldots A[p-1]$, where each array entry is either 0 or 1. Moreover, assume that at most half of the array entries are 1.

Suppose we chose $a, b \in \{0, \ldots, p-1\}$ at random, and then we query k array entries,

$$A[(a + b) \bmod p], A[(2a + b) \bmod p], A[(3a + b) \bmod p, \ldots, A[(ka + b) \bmod p].$$

(a) Use Chebyshev's Inequality to determine an upper bound for the probability that we find no array entry with value 1.

(b) Now assume that you query k locations in the array chosen uniformly and independently at random. What is now an upper bound for the probability that you find no 1?

(c) Discuss how many random bits are needed for each of the two methods above?

(d) Discuss how the first method of generating randomness could be used to reduce the error probability of certain Monte Carlo algorithms, which either give a "true" or "false" answer, and allow only false negatives but not false positives.

probability space, average case analysis

2.23 Consider a list $L[1], \ldots, L[n]$ of values. Let v be a value that may or may not occur in the list. Moreover, the probability that v is the value of $L[i]$ is exactly $i/(n(n+1))$.

[2] See Appendix 2.5 for some hints.

(a) What is the probability that v does not occur in the list?

(b) How many iterations does the linear search algorithm presented in class need in expectation, until it either finds v, or decides that v is not in the list?

Monte Carlo algorithm

2.24 Devise a Monte Carlo algorithm that determines whether a given array of n integers is sorted. A step of the algorithm should answer "true" if it determines the list is not sorted, and "unknown" otherwise. After k steps, the algorithm decides that the integers are sorted, if the answer is "unknown" in each step. Otherwise, the algorithm decides that the integers are not sorted.

What are the probabilities of false negative and false positive decisions, respectively?

2.5. Selected Solutions

Exercise 2.9

(a) In each query, the algorithm chooses an array entry uniformly at random among all array entries. Since half of the array entries have value 1, it queries a 1 with probability 1/2. This is also true for the first query, so the algorithm stops after the first query with probability 1/2.

(b) Let X be the number of queries until the algorithm stops. For the algorithm to stop after exactly k queries, it must first query $k-1$ 0s and then a 1. Each of the events, querying a 0, and querying a 1 has probability 1/2. Hence, the probability that it stops after exactly k queries is $(1/2)^{k-1} \cdot (1/2) = (1/2)^k$. In other words, $p(X = k) = (1/2)^k$.

(c) X is a geometrically distributed random variable with parameter 1/2. (Define "success" as querying a 1. Then each query has a success probability of 1/2, and X is the number of queries until success.) Thus, $E[X] = 1/(1/2) = 2$ by Theorem 2.15.

Exercise 2.10

(a) If we roll three dice, the probability that the sum of all pips is even, is 1/2. We can see this as follows: Fix the first two dice. If their sum of pips is even, the last die roll will yield an even sum, if it is also even. This happens with probability 1/2. If the sum of the pips of the first two dice is odd, then the last die roll must also be odd to yield an overall even sum. Again, this happens with probability 1/2.

Now let X denote the number of times we have to repeat 3 die rolls, until the sum of pips is even. Then X is geometrically distributed with parameter 1/2. Hence, $E[X] = 2$.

(b) Let us call it success, when the three dice show an even number of pips for the first time. We know from part (a) that the probability of success is 1/2. Let X be the number of times we roll three dice until we have success, or until we have rolled 10 times. Then for $k \in \{0, \ldots, 9\}$ we have $X > k$ if and only if the first k experiments fail. This happens with probability $1/2^k$. Hence, we have

$$p(X > k) = (1/2)^k \quad \text{for } k = 0, \ldots, 9.$$

Moreover, since we stop after performing the experiment 10 times, we have

$$p(X > 10) = 0.$$

Then

$$E[X] = \sum_{k=0}^{\infty} p(X > k) = \sum_{k=0}^{9} \frac{1}{2^k} + \sum_{k=10}^{\infty} 0 = \sum_{k=0}^{9} \left(\frac{1}{2}\right)^k$$
$$= \frac{1 - 1/2^{10}}{1 - 1/2} = 2 - \frac{1}{2^9}.$$

Exercise 2.11

(a) We repeat algorithm A n times. If A outputs "composite" only once in those n rounds, then we also output "composite". Otherwise, we output "prime".

Let x be the input. First assume that x is a prime. By the assumptions, A cannot output "composite" if the input is prime, so all n runs of A must output prime. Hence, our new algorithm also correctly outputs prime.

Now assume that the input x is composite. Then our new algorithm makes an error, if A outputs "prime" n times. In each run this happens with probability at most $1 - 1/n$. So the probability that this happens in each of the n runs is at most

$$(1 - 1/n)^n \leq 1/e < 1/2.$$

Hence, the error probability of our new algorithm is less than $1/2$.

(b) We repeat algorithm A (by running it on the input x) until it outputs "prime" or "composite". The pseudocode for this is as follows:

(a) **Repeat**
Run A on input x, and let *out* be the output.
until *out* $\neq ?$

(b) Output *out*.

Let X be the random variable that counts the number of times we run A on input x. Each time we run A, the probability that A outputs "prime" or "composite" is $q \geq 1/n$. Hence, X is geometrically distributed with parameter q. Therefore, $E[X] = 1/q \leq \frac{1}{1/n} = n$.

Each time we run A on input x, the algorithm performs at most $T(n)$ steps. Hence, the total running time (when running algorithm A X times) is

$$E[X \cdot T(n)] = E[X] \cdot T(n) \leq n \cdot T(n).$$

Exercise 2.13 We count the number of integers generated until we see the first 1, and let that number be X_1. After we have seen the first 1, we count again, until we see the first 2, and let the count be X_2. Thus, now we have seen a 1 and after that a 2. We count again, until we see the first 3, and we let the number of integers generated be X_3. And so on.

More precisely, once the random integers contain a subsequence $(1, \ldots, i-1)$ for the first time, X_i denotes the number of additional integers generated, until we see the first i, and thus the resulting sequence contains $(1, \ldots, i)$ for the first time. Then $X_1 + \cdots + X_i$ is the number of integers generated, until we have obtained a sequence that contains $(1, \ldots, i)$ as a subsequence. So for $X = X_1 + \cdots + X_n$, we need to compute $E[X]$.

For example, consider the following random sequence, and the corresponding values of X_i for $n = 4$:

$$\underbrace{\underbrace{2, 3, 4, 4, 2, \mathbf{1}}_{X_1=6}, \underbrace{3, 1, \mathbf{2}}_{X_2=3}, \underbrace{2, 4, 1, 4, \mathbf{3}}_{X_3=5}, \underbrace{1, 2, 2, 1, 3, 1, 3, 2, \mathbf{4}}_{X_4=9}}_{X=X_1+X_2+X_3+X_4=23}.$$

Each random variable X_i counts the number of integers generated, until value i occurs. Thus, defining the generation of i as "success", we count the number of repetitions of a random experiment until success occurs, and the success probability is $1/n$. Hence, X_i is geometrically distributed with parameter $1/n$, and thus has expectation $E[X_i] = n$ (by Theorem 2.15). By linearity of expectation,

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = n \cdot n = n^2.$$

Exercise 2.14

(a) Let Y denote the number of rounds until one of the players wins. Let us define a round as successful, if there is no tie in that round. The probability that that happens is $5/6$ (we can fix Alice's die result, and then for 5 out of 6 choices, Bob's die roll will be different). Hence, Y is geometrically distributed with parameter $5/6$, and so $E[X] = 6/5$ (by Theorem 2.15).

Now let X be the total number of die rolls. Since Y is the total number of rounds, and each round comprises 2 die rolls, we have $X = 2Y$. Thus, by linearity of expectation

$$E[X] = E[2Y] = 2E[Y] = 2 \cdot \frac{6}{5} = \frac{12}{5}.$$

(b) For a given round, let T denote the event that there is a tie, and let A denote the event that Alice wins the round. In part (a) we already determined $p\big(\overline{T}\big) = 5/6$. Now observe that if there is no tie, then each of the two players is equally likely to win, so $p\big(A \bigm| \overline{T}\big) = 1/2$. Thus, the probability that Alice wins a round is

$$p(A) = p\big(A \cap \overline{T}\big) = p\big(A \bigm| \overline{T}\big) \cdot p\big(\overline{T}\big) = \frac{1}{2} \cdot \frac{5}{6} = \frac{5}{12}.$$

Let Z_1 denote the number of rounds until Alice has won her first round. Similarly, let Z_i denote the number of rounds they need to play after Alice has won $i-1$ rounds, until she wins for the i-th time. Then Z_i is geometrically distributed with parameter $5/12$, and thus $E[Z_i] = 12/5$ (by Theorem 2.15).

If Z is the total number of rounds until Alice has won 5 times, then $Z = Z_1 + \cdots + Z_5$, and so by linearity of expectation

$$E[Z] = E[Z_1 + \cdots + Z_5] = E[Z_1] + \cdots + E[Z_5] = 5 \cdot \frac{12}{5} = 12.$$

Exercise 2.22 (a) Here are some hints:

- Let $y_i = (ia + b) \bmod p$ for random $a, b \in \{0, \ldots, p-1\}$. The algorithm queries $A[y_1], A[y_2], \ldots, A[y_k]$. Let X_i be the indicator random variable that has the same value as $A[y_i]$. As stated in the question, the random variables $y_1, \ldots, y_k$ are pairwise independent, and thus so are the random variables $X_1, \ldots, X_k$.
- Let X denote the *number* of queries of array entries with value 1. Then part (a) asks you to determine an upper bound for $p(X = 0)$. Chebyshev's Inequality can help with that.

A. Selected Identities, Inequalities, and Theorems

A.1. Sums and Bounds

Let $x \geq 1$ and $a \neq 1$. Then

$$(1 - 1/x)^x \leq 1/e$$

$$\ln(n+1) \leq \sum_{i=1}^{n} \frac{1}{i} \leq (\ln n) + 1 \qquad \text{(Harmonic Series)}$$

$$\sum_{i=0}^{k} a^i = \frac{1 - a^{k+1}}{1 - a} \qquad \text{(finite Geometric Series)}$$

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1 - a} \quad \text{if } |a| < 1 \qquad \text{(infinite Geometric Series)}$$

A.2. Probabilities of Events

Let A and B be events of the sample space Ω, and let $p : \Omega \to [0, 1]$ be a probability distribution.

$$\sum_{w \in \Omega} p(w) = 1 \qquad \text{(axiom of a probability space)}$$

$$p(A) = \sum_{w \in A} p(w) \qquad \text{(probability of an event)}$$

$$p\big(\overline{A}\big) = 1 - p(A) \qquad \text{(complement)}$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \qquad \text{(inclusion-exclusion)}$$

$$p(A_1 \cup A_2 \cup \cdots \cup A_k) \leq p(A_1) + \cdots + p(A_k) \qquad \text{(union bound)}$$

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} \qquad \text{(conditional probability)}$$

$$p(A) = p(A \mid B)p(B) + p\big(A \mid \overline{B}\big)p\big(\overline{B}\big) \qquad \text{(total probability theorem)}$$

$$p(A \cap B) = p(A \mid B) \cdot p(B) \qquad \text{(intersection)}$$

$$p(A \cap B) = p(A) \cdot p(B) \quad \Leftrightarrow \quad A \text{ and } B \text{ are independent} \qquad \text{(independence)}$$

$$p(A \mid B) = \frac{p(B \mid A) \cdot p(A)}{p(B)} \qquad \text{(Bayes' Theorem)}$$

A.3. Random Variables

Let $X : \Omega \to V_X$ and $Y : \Omega \to V_Y$ be random variables, and $a, t \in \mathbb{R}$.

$$E[X] = \sum_{w \in \Omega} p(w) \cdot X(w) = \sum_{i \in V_X} i \cdot p(X = i) \qquad \text{(expectation)}$$

$$E[X] = \sum_{i=0}^{\infty} p(X > i), \quad \text{if } V_X = \mathbb{N}_0 \qquad \text{(expectation of a non-negative integer RV)}$$

$$E[a \cdot X] = a \cdot E[X] \qquad \text{(linearity of expectation)}$$

$$E[X + Y] = E[X] + E[Y] \qquad \text{(linearity of expectation)}$$

$$\text{Var}[X] = E\big[(X - E[X])^2\big] = E[X^2] - E[X]^2 \qquad \text{(variance)}$$

$$p(X \geq t) \leq \frac{E[X]}{t}, \quad \text{if } t > 0 \text{ and } V_X \subseteq \mathbb{R}_{\geq 0} \qquad \text{(Markov's Inequality)}$$

$$p(|X - E[X]| \geq t) \leq \frac{\text{Var}[X]}{t^2} \quad \text{if } t > 0 \qquad \text{(Chebyshev's Inequality)}$$

A.4. Geometric Distribution

Let X be a geometrically distributed random variable with parameter q, and let $k \in \mathbb{N}_0$.

$$p(X = k) = (1 - q)^{k-1} q, \text{ if } k > 0 \qquad \text{(distribution)}$$

$$p(X > k) = (1 - q)^k$$

$$p\Big(X \geq \big\lceil cE[X] \big\rceil\Big) < \left(\frac{1}{e}\right)^c, \text{ if } c \geq 1 \qquad \text{(tail bound)}$$

$$E[X] = \frac{1}{q} \qquad \text{(expectation)}$$