Queen's University
School of Computing

CISC 203: Discrete Mathematics for Computing II
Module 8: Probability
Fall 2021

This module corresponds to the following sections from your textbook:

30. Sample Space
31. Events

1 Probability Theory and Computing

Fundamentally, a computer is a deterministic machine. By "deterministic", we mean that if we perform the same actions multiple times, we will produce the same result each time. Each time we run a (simple, sequential) program, the CPU executes the same instructions and the same values are stored in memory. Indeed, if this were not the case, then computers would not be a very useful tool at all. Why, then, do computer scientists care about probability theory; a field where performing the same action multiple times can result in multiple, distinct outcomes occurring with varying odds?

Well, for starters, not all programs are sequential, as we assumed previously. Some programs are parallelized (e.g., multithreaded or multiprocess), with different parts of the program running simultaneously and possibly finishing their computations in a different order each time the program is run. Other programs are randomized, so that each run of the program results in a potentially different output. In fact, some programs rely on parallelization or randomization to produce output in a reasonable amount of time (or at all). Thus, it is important for computer scientists to have a working knowledge of probability theory in order to understand more advanced topics like parallel computing and randomized algorithms.

There are two branches of probability theory: discrete probability and continuous probability. Discrete probability is probability theory in the context of finite or countable probability spaces, where we can compute probabilities of multiple events via summation. Continuous probability, on the other hand, is probability theory in the context of uncountable probability spaces, where we can compute probabilities of multiple events via integration. In keeping with the theme of this course, we will focus here only on discrete probability; continuous probability is a whole other beast, and discussion on that topic will be left for another course.

2 Sample Space

Probability theory is based on the idea of observing the outcome of an experiment that has different possible outcomes. For instance, a coin flip experiment can result in two possible outcomes: heads or tails. Similarly, the roll of a six-sided die can result in six possible outcomes, and drawing a random card from a standard deck of playing cards can result in 52 possible outcomes. Each outcome has a probability of occurring—in the experiments that we mentioned so far, all outcomes are equally probable. Probabilities are expressed as real numbers between 0 and 1. An event with probability 1 is certain to occur, and an event with probability 0 is impossible. For any given experiment, the sum of the probabilities of all the possible outcome should equal 1—this follows from the fact that performing an experiment must certainly result in an observable outcome.

Definition 1 (Sample space)**.** A **sample space** is a pair (S, P) where S is a finite, nonempty set of outcomes and $P : S \longrightarrow \mathbb{R}$ is a function such that $P(s) \geq 0$ for all $s \in S$ and

$$\sum_{s \in S} P(s) = 1.$$

A sample space is also called a **probability space**.

Example 2. The sample space for the experiment of flipping a single coin consists of the set of two outcomes

$$S_{\mathrm{C}} = \{Ⓗ, Ⓣ\},$$

and the probability function $P_C : S_C \longrightarrow \mathbb{R}$ defined by

$$P_C(Ⓗ) = \frac{1}{2} \qquad\qquad P_C(Ⓣ) = \frac{1}{2}.$$

Example 3. The sample space for the experiment of rolling a single die consists of the set of six outcomes

$$S_{\mathrm{D}} = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\},$$

and the probability function $P_D : S_D \longrightarrow \mathbb{R}$ defined by

$$P_D(⚀) = \frac{1}{6} \qquad P_D(⚁) = \frac{1}{6} \qquad P_D(⚂) = \frac{1}{6} \qquad P_D(⚃) = \frac{1}{6} \qquad P_D(⚄) = \frac{1}{6} \qquad P_D(⚅) = \frac{1}{6}.$$

This probability function assumes, of course, that we are using a fair die, as opposed to rigged or unfair dice that might behave differently.

Example 4. The sample space for the experiment of drawing a single playing card from a deck consists the set of the 52 outcomes

$$S_{\mathrm{PC}} = \left\{ \begin{array}{ccccccccccccc} \mathsf{A}\clubsuit, & \mathsf{K}\clubsuit, & \mathsf{Q}\clubsuit, & \mathsf{J}\clubsuit, & \mathsf{10}\clubsuit, & \mathsf{9}\clubsuit, & \mathsf{8}\clubsuit, & \mathsf{7}\clubsuit, & \mathsf{6}\clubsuit, & \mathsf{5}\clubsuit, & \mathsf{4}\clubsuit, & \mathsf{3}\clubsuit, & \mathsf{2}\clubsuit \\ \mathsf{A}\diamondsuit, & \mathsf{K}\diamondsuit, & \mathsf{Q}\diamondsuit, & \mathsf{J}\diamondsuit, & \mathsf{10}\diamondsuit, & \mathsf{9}\diamondsuit, & \mathsf{8}\diamondsuit, & \mathsf{7}\diamondsuit, & \mathsf{6}\diamondsuit, & \mathsf{5}\diamondsuit, & \mathsf{4}\diamondsuit, & \mathsf{3}\diamondsuit, & \mathsf{2}\diamondsuit \\ \mathsf{A}\heartsuit, & \mathsf{K}\heartsuit, & \mathsf{Q}\heartsuit, & \mathsf{J}\heartsuit, & \mathsf{10}\heartsuit, & \mathsf{9}\heartsuit, & \mathsf{8}\heartsuit, & \mathsf{7}\heartsuit, & \mathsf{6}\heartsuit, & \mathsf{5}\heartsuit, & \mathsf{4}\heartsuit, & \mathsf{3}\heartsuit, & \mathsf{2}\heartsuit \\ \mathsf{A}\spadesuit, & \mathsf{K}\spadesuit, & \mathsf{Q}\spadesuit, & \mathsf{J}\spadesuit, & \mathsf{10}\spadesuit, & \mathsf{9}\spadesuit, & \mathsf{8}\spadesuit, & \mathsf{7}\spadesuit, & \mathsf{6}\spadesuit, & \mathsf{5}\spadesuit, & \mathsf{4}\spadesuit, & \mathsf{3}\spadesuit, & \mathsf{2}\spadesuit \end{array} \right\},$$

and the probability function $P_{PC} : S_{PC} \longrightarrow \mathbb{R}$ defined by $P(s) = \frac{1}{52}$ for all $s \in S_{PC}$.

Example 5. Consider a box containing two red balls and one blue ball. If we reach into the box and take out one ball, there are two possible outcomes $S_B = \{🔴, 🔵\}$.

Assuming that each ball has an identical size and weight, and may differ only in colour, the probability function $P_B : S_B \longrightarrow \mathbb{R}$ is defined by

$$P_B(🔴) = \frac{2}{3} \qquad\qquad P_B(🔵) = \frac{1}{3}.$$

The above example illustrates that the outcomes of an experiment do not need to be equally probable.

3 Events

We use the word **event** to describe any subset of S from a sample space (S, P). For example, the event $A = \{⚁, ⚃, ⚅\}$ is the subset of S_D that includes all outcomes where the roll of a die is an even number. Thus, if we roll a die and the result is an even number such as ⚁, we say that A **has occurred**.

Definition 6 (Probability of an Event)**.** Let (S, P) be a sample space. An event A is a subset of S, i.e., $A \subseteq S$. The probability of the event A occurring is

$$P(A) = \sum_{a \in A} P(a).$$

Example 7. Recall the “rolling dice” experiment, with $S_{\mathrm{D}} = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$. If we instead roll two dice, there are 36 possible outcomes defined by

$$S_{\mathrm{2D}} = S_{\mathrm{D}} \times S_{\mathrm{D}} = \{\text{⚀⚀}, \text{⚀⚁}, \text{⚀⚂}, \ldots, \text{⚅⚄}, \text{⚅⚅}\}.$$

Each outcome is equally likely, and thus $P_{2D}(s) = 1/36$ for all $s \in S_{2D}$.

What is the probability of rolling two dice and obtaining a sum of 5? Let A denote the event that the numbers on the dice sum to 5. So, we have

$$A = \{\text{⚀⚃}, \text{⚁⚂}, \text{⚂⚁}, \text{⚃⚀}\}.$$

The probability of A occurring is

$$\begin{aligned} P(A) &= P(\text{⚀⚃}) + P(\text{⚁⚂}) + P(\text{⚂⚁}) + P(\text{⚃⚀}) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{9}. \end{aligned}$$

Exercise

What is the probability of rolling a pair of dice where the sum of the numbers on the two dice is 7? Compare your answer with Example 31.2 on p. 217 of the textbook.

Example 8. Recall the “flipping coins” experiment, with $S_{\mathrm{C}} = \{\mathrm{H}, \mathrm{T}\}$. If we instead flip three coins, there are eight possible outcomes defined by

$$S_{\mathrm{3C}} = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}.$$

Since each outcome is equally likely, we have $P_{3C}(s) = 1/8$ for all $s \in S_{3C}$.

What is the probability that exactly two H are shown? Let B be the event that exactly two coin flips show H. Thus,

$$B = \{\mathrm{HHT}, \mathrm{THH}, \mathrm{HTH}\}.$$

So, the probability of B occurring is

$$\begin{aligned} P(B) &= P(\mathrm{HHT}) + P(\mathrm{THH}) + P(\mathrm{HTH}) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}. \end{aligned}$$

However, we did not actually need to write out the elements of B explicitly. We know that each event in S_{3C} is equally probable, and $B \subseteq S_{3C}$, and $|B|$ is the number of ways that two out of three coins can show H. So, we have

$$|B| = \binom{3}{2} = 3,$$

and thus

$$P(B) = |B| \cdot \frac{1}{8} = \frac{3}{8}.$$

Exercise

Express the probability of flipping five coins, with exactly two of them showing H, in terms of a binomial coefficient as in the example above. Then, evaluate the result and compare with Example 31.3 on p. 218 of the textbook.

Example 9. Recall the "drawing a card" experiment, with the sample space S_{PC} and probability function $P(s) = \frac{1}{52}$ for all $s \in S_{PC}$.

In a game of poker, each player is dealt a "hand" of five cards. The sample space S_{PC5} is defined by the set where each outcome $h \in S_{\mathrm{PC5}}$ represents a hand of five cards. So, we have $|S_{\mathrm{PC5}}| = \binom{52}{5}$. Since each outcome is equally probable, every poker hand has a probability $P(h) = 1/\binom{52}{5}$ of occurring.

A "full house" is a poker hand that consists of five cards: three cards are of one value (e.g., three Queens), and the other two cards are of another value (e.g., two 7s). For example, Q♦ Q♥ Q♠ 7♣ 7♥ is a full house. What is the probability that h is a full house?

Let A be the event that the poker hand is a full house. Since every poker hand has a probability of $1/\binom{52}{5}$, the probability of a full house is $|A|/\binom{52}{5}$.

To determine $|A|$, we must calculate the number of ways we can draw a full house. First, we choose the value of the first three cards: we can do this in $\binom{13}{1}$ ways. Then, we choose three out of the four suits of that value in $\binom{4}{3}$ ways. Next, we choose the value of other two cards in $\binom{12}{1}$ ways, and the suits of that value in $\binom{4}{2}$ ways. The total number of ways of drawing a full house is thus

$$|A| = \binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{1} \cdot \binom{4}{2}$$

and the probability of a full house occurring is

$$|A| \cdot \frac{1}{\binom{52}{5}} = \frac{\binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{1} \cdot \binom{4}{2}}{\binom{52}{5}} = \frac{(13)(4)(12)(6)}{2\,598\,960} = 0.00144.$$

Exercise

A poker hand is called "four of a kind" if four of the five cards have the same value (e.g., all 7s or all Queens). What is the probability that a poker hand is four of a kind? Compare your answer with Example 31.5 on p. 218 in the textbook.

3.1 Combining Events

Recall from Definition 6 that an event is a subset of a sample space. So, we can use set operations to combine events.

Definition 10 (Union of events)**.** If A and B are events, $A \cup B$ is the event where either A or B, or both, occur.

Definition 11 (Intersection of events)**.** If A and B are events, $A \cap B$ is the event where both A and B occur.

Definition 12 (Difference of events)**.** If A and B are events, $A - B$ is the event that A occurs but B does not.

Definition 13 (Complement of events)**.** The event $\overline{A}$ is the event where A does not occur. For a sample space (S, P), we have $\overline{A} = S - A$, since S is the set of all possible outcomes.

Example 14. Recall the "rolling a die" experiment, with $S_{\mathrm{D}} = \{⚀,⚁,⚂,⚃,⚄,⚅\}$ and $P(s) = \frac{1}{6}$ for all $s \in S_D$. Let $A = \{⚁,⚃,⚅\}$ be the event that a die shows an even number, and $B = \{⚁,⚂,⚄\}$ be the event that the die shows a prime number.

$A \cup B$ is the event that the die shows a number that is even or prime (or both). So we have $A \cup B = \{⚁,⚂,⚃,⚄,⚅\}$ and the probability of the event $A \cup B$ occurring is

$$P(A \cup B) = |A \cup B| \cdot \frac{1}{6} = \frac{5}{6}.$$

$A \cap B$ is the event that the die shows a number that is both even and prime. So, we have $A \cap B = \{⚁\}$ and the probability of $A \cap B$ occurring is

$$P(A \cap B) = |A \cap B| \cdot \frac{1}{6} = \frac{1}{6}.$$

$A - B$ is the event that the die shows a number that is even but not prime. So, we have $A - B = \{⚃,⚅\}$ and the probability of $A - B$ occurring is

$$P(A - B) = |A - B| \cdot \frac{1}{6} = \frac{2}{6}.$$

$\overline{A}$ is the event that the die does not show an even number. So, we have $\overline{A} = S_{\mathrm{D}} - \{⚁,⚃,⚅\} = \{⚀,⚂,⚄\}$ and the probability of $\overline{A}$ occurring is

$$P(\overline{A}) = |\overline{A}| \cdot \frac{1}{6} = \frac{3}{6}.$$

3.2 Mutually Exclusive Events

Is it possible to find $P(A \cup B)$ if we are given only $P(A)$ and $P(B)$? The answer is no, as illustrated by the following example.

Example 15. Consider again the "rolling a die" experiment, with $S_{\mathrm{D}} = \{⚀,⚁,⚂,⚃,⚄,⚅\}$ and $P(s) = \frac{1}{6}$ for all $s \in S_D$.

Let A be the event where an even number is rolled, and the event B be the event that a prime number is rolled. Then, we have $P(A) = P(B) = \frac{1}{2}$ and $P(A \cup B) = \frac{5}{6}$.

Let C be the event where an odd number is rolled. Then, we have $P(A) = P(C) = \frac{1}{2}$ and $P(A \cup C) = 1$.

The example above should remind you of the Inclusion-Exclusion Principle that we studied earlier in the course. The following proposition demonstrates the analogous property that we have in probability theory.

Proposition 16. *Let A and B be events in a sample space (S, P). Then,*

$$P(A) + P(B) = P(A \cup B) + P(A \cap B).$$

Proof. This can be done as a direct proof: See Proposition 31.7 on p. 219 of your textbook. □

Example 17. From Example 14 above, note that we have

$$P(A) = 1/2 \qquad P(B) = 1/2 \qquad P(A \cup B) = 5/6 \qquad P(A \cap B) = 1/6.$$

So, we see that $P(A) + P(B) = P(A \cup B) + P(A \cap B)$.

If A and B are disjoint (i.e., $A \cap B = \emptyset$), we call them **mutually exclusive**, and it follows from the proposition above that $P(A \cup B) = P(A) + P(B)$ for such events.

Exercise

Show, via a direct proof using Proposition 16 and the fact that all probabilities must be nonnegative, that

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

This is a relatively trivial one-liner proof. Boole's Inequality generalizes this property to any finite or countable set of events.

Let us now consider: Is it possible to find $P(\overline{A})$ if we are given only $P(A)$? The answer is yes, as we show below.

Theorem 18. *Let A be an event occurring with probability $P(A)$. The probability that the event $\overline{A}$ occurs is $P(\overline{A}) = 1 - P(A)$.*

Proof. By Definition 1, the sum of the probabilities of all outcomes in S must equal 1. Since each outcome is either in A or $\overline{A}$, but not in both, we have

$$\sum_{s \in S} P(s) = 1 = P(A) + P(\overline{A}).$$

So, we have $P(\overline{A}) = 1 - P(A)$. □

Case Study: Birthday Problem

The well-known "Birthday problem" in probability theory considers the probability of two or more people having the same birthday in a group of n randomly-chosen people. For $n = 4$, the probability is only 1.64%. For $n = 23$, the probability is 50.73%. For $n = 367$, the probability reaches 100% due to the Pigeonhole Principle that we recently studied.

The textbook calculates the above probabilities on p. 220 using the above property from Theorem 18. You may also be interested in checking out the introduction in the Wikipedia article on the Birthday problem. One of the real-world computing applications of the Birthday problem is a cryptographic attack called the birthday attack.

Example 19. Let b be a randomly-generated bit sequence of length four, and define events $E =$ "b contains all `1`s" and $F =$ "b starts with two `0`s". Assuming that each bit in the sequence is equally likely to be generated as a `0` or `1`, we have $P(E) = {}^1/_{16}$ and $P(F) = {}^1/_4$.

What is $\overline{E}$? This is the event where b does not contain all `1`s; that is, where b contains at least one `0`. By Theorem 18, we have that $P(\overline{E}) = 1 - P(E) = {}^{15}/_{16}$. This makes sense, since there are 15 bit strings of length four containing at least one `0`.

What is $P(E \cap F)$? This is the event where the first bit of b is simultaneously `1` and `0`; this is obviously impossible, since the bit string b cannot begin with both a `1` and a `0`. Therefore, E and F are mutually exclusive and $P(E \cap F) = 0$.

What is $P(E \cup F)$? This is the event where b either contains all `1`s or starts with two `0`s. Since these events are mutually exclusive, we have $P(E \cup F) = P(E) + P(F) = {}^1/_{16} + {}^1/_4 = {}^5/_{16}$.