1

BASIC CONCEPTS

1.1 DEFINITIONS AND PROPERTIES

Most of the problems in probability and statistics involve determining how likely it is that certain things will occur. Before we can talk about what is likely or unlikely, we need to know what is possible. In other words, we need some framework in which to discuss what sorts of things have the potential to occur. To that end, we begin by introducing the basic concepts of "sample space", "experiment", "outcome", and "event". We also define what we mean by a "probability" and provide some examples to demonstrate the consequences of the definition.

1.1.1 *Definitions*

DEFINITION 1.1.1. **(Sample Space)** *A sample space S is a set. The elements of the set S will be called "outcomes" and should be viewed as a listing of all possibilities that might occur. We will call the process of actually selecting one of these outcomes an "experiment".*

For its familiarity and simplicity we will frequently use the example of rolling a die. In that case our sample space would be $S = \{1, 2, 3, 4, 5, 6\}$, a complete listing of all of the outcomes on the die. Performing an experiment in this case would mean rolling the die and recording the number that it shows. However, sample space outcomes need not be numeric. If we are flipping a coin (another simple and familiar example) experiments would result in one of two outcomes and the appropriate sample space would be $S = \{Heads, Tails\}$.

For a more interesting example, if we are discussing which country will win the next World Cup, outcomes might include Brazil, Spain, Canada, and Thailand. Here the set S might be all the world's countries. An experiment in this case requires waiting for the next World Cup and identifying the country which wins the championship game. Though we have not yet explained how probability relates to a sample space, soccer fans amongst our readers may regard this example as a foreshadowing that not all outcomes of a sample space will necessarily have the same associated probabilities.

DEFINITION 1.1.2. **(Temporary Definition of Event)** *Given a sample space S, an "event" is any subset $E \subset S$.*

This definition will allow us to talk about how likely it is that a range of possible outcomes might occur. Continuing our examples above we might want to talk about the probability that a die rolls a number larger than two. This would involve the event $\{3, 4, 5, 6\}$ as a subset of $\{1, 2, 3, 4, 5, 6\}$. In the soccer example we might ask whether the World Cup will be won by a South American country. This subset of our list of all the world's nations would contain Brazil as an element, but not Spain.

It is worth noting that the definition of "event" includes both S, the sample space itself, and $\varnothing$, the empty set, as legitimate examples. As we introduce more complicated examples we will see that it is not always necessary (or even possible) to regard every single subset of a sample space as a legitimate event, but since the reasons for that may be distracting at this point we will use the above as a temporary definition of "event" and refine the definition when it becomes necessary.

To each event, we want to assign a chance (or "probability") which will be a number between 0 and 1. So if the probability of an event E is 0.72, we interpret that as saying, "When an experiment is performed, it has a 72% chance of resulting in an outcome contained in the event E". Probabilities will satisfy two axioms stated and explained below. This formal definition is due to Andrey Kolmogorov (1903-1987).

Definition 1.1.3. **(Probability Space Axioms)** *Let S be a sample space and let $\mathcal{F}$ be the collection of all events.*
*A "**probability**" is a function $P : \mathcal{F} \to [0, 1]$ such that*

(1) $P(S) = 1$; and

(2) If $E_1, E_2, \ldots$ are a countable collection of disjoint events (that is, $E_i \cap E_j = \varnothing$ if $i \neq j$), then

$$P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j). \tag{1.1.1}$$

The first axiom is relatively straight forward. It simply reiterates that S did, indeed, include all possibilities, and therefore there is a 100% chance that an experiment will result in some outcome included in S. The second axiom is not as complicated as it looks. It simply says that probabilities add when combining a countable number of disjoint events. It is implicit that the series on right hand side of the equation (1.1.1) converges. Further (1.1.1) also holds when combining finite number of disjoint events (see Theorem 1.1.4 below).

Returning to our World Cup example, suppose A is a list of all North American countries and E is a list of all European countries. If it happens that $P(A) = 0.05$ and $P(E) = 0.57$ then $P(A \cup E) = 0.62$. In other words, if there is a 5% chance the next World Cup will be won by a North American nation and a 57% chance that it will be won by a European nation, then there is a 62% chance that it will be won by a nation from either Europe or North America. The disjointness of these events is obvious since (if we discount island territories) there isn't any country that is in both North America and Europe.

The requirement of axiom two that the collection of events be countable is important. We shall see shortly that, as a consequence of axiom two, disjoint additivity also applies to any finite collection of events. It does not apply to uncountably infinite collections of events, though that fact will not be relevant until later in the text when we discuss continuous probability spaces.

1.1.2 *Basic Properties*

There are some immediate consequences of these probability axioms which we will state and prove before returning to some simple examples.

THEOREM 1.1.4. *Let P be a probability on a sample space S. Then,*

(1) $P(\varnothing) = 0$;

(2) If $E_1, E_2, ...E_n$ are a finite collection of disjoint events, then

$$P(\bigcup_{j=1}^{n} E_j) = \sum_{j=1}^{n} P(E_j);$$

(3) If E and F are events with $E \subset F$, then $P(E) \le P(F)$;

(4) If E and F are events with $E \subset F$, then $P(F \setminus E) = P(F) - P(E)$;

(5) Let E^c be the complement of event E. Then $P(E^c) = 1 - P(E)$; and

(6) If E and F are events then $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Proof of (1) - The empty set is disjoint from itself, so $\varnothing, \varnothing, \ldots$ is a countable disjoint collection of events. From the second axiom, $P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$. When this is applied to the collection of empty sets we have $P(\varnothing) = \sum_{j=1}^{\infty} P(\varnothing)$. If $P(\varnothing)$ had any non-zero value, the right hand side of this equation would be a divergent series while the left hand side would be a number. Therefore, $P(\varnothing) = 0$.

Proof of (2) - To use axiom two we need to make this a countable collection of events. We may do so while preserving disjointness by including copies of the empty set. Define $E_j = \varnothing$ for $j > n$. Then $E_1, E_2, \ldots, E_n, \varnothing, \varnothing, \ldots$ is a countable collection of disjoint sets and therefore $P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$. However, the empty sets add nothing to the union and so $\bigcup_{j=1}^{\infty} E_j = \bigcup_{j=1}^{n} E_j$. Likewise since we have shown $P(\varnothing) = 0$ these sets also add nothing to the sum, so $\sum_{j=1}^{\infty} P(E_j) = \sum_{j=1}^{n} P(E_j)$.

Combining these gives the result:

$$P(\bigcup_{j=1}^{n} E_j) = P(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j) = \sum_{j=1}^{n} P(E_j).$$

Proof of (3) - If $E \subset F$, then E and $F \setminus E$ are disjoint events with a union equal to F. Using (2) above gives $P(F) = P(E \cup (F \setminus E)) = P(E) + P(F \setminus E)$.
Since probabilities are assumed to be positive, it follows that $P(F) \ge P(E)$.

Proof of (4) - As with the proof of (3) above, E and $F \setminus E$ are disjoint events with $E \cup (F \setminus E) = F$. Therefore $P(F) = P(E) + P(F \setminus E)$ from which we get the result.

Proof of (5) - This is simple a special case of (4) where $F = S$.

Proof of (6) - We may disassemble $E \cup F$ disjointly as $E \cup F = E \cup (F \setminus E)$. Then from (2) we have $P(E \cup F) = P(E) + P(F \setminus E)$.
Next, since $F \setminus E \subset F$ and since $F \setminus (F \setminus E) = E \cap F$ we can use (4) to write $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. ■

Example 1.1.5. A coin flip can come up either "heads" or "tails", so $S = \{heads, tails\}$. A coin is considered "fair" if each of these outcomes is equally likely. Which axioms or properties above can be used to reach the (obvious) conclusion that both outcomes have a 50% chance of occurring?

Each outcome can also be regarded as an event. So $E = \{heads\}$ and $F = \{tails\}$ are two disjoint events. If the coin is fair, each of these events is equally likely, so $P(E) = P(F) = p$ for some value of p. However, using the second axiom, $1 = P(S) = P(E \cup F) = P(E) + P(F) = 2p$. Therefore, $p = 0.5$, or in other words each of the two possibilities has a 50% chance of occurring on any flip of a fair coin. ■

In the examples above we have explicitly described the sample space S, but in many cases this is neither necessary nor desirable. We may still use the probability space axioms and their consequences when we know the probabilities of certain events even if the sample space is not explicitly described.

Example 1.1.6. A certain sea-side town has a small fishing industry. The quantity of fish caught by the town in a given year is variable, but we know there is a 35% chance that the town's fleet will catch over 400 tons of fish, but only a 10% chance that they will catch over 500 tons of fish. How likely is it they will catch between 400 and 500 tons of fish?

The answer to this may be obvious without resorting to sets, but we use it as a first example to illustrate the proper use of events. Note, though, that we will not explicitly describe the sample space S.

There are two relevant events described in the problem above. We have F representing "the town's fleet will catch over 400 tons of fish" and E representing "the town's fleet will catch over 500 tons of fish". We are given that $P(E) = 0.1$ while $P(F) = 0.35$.

Of course $E \subset F$ since if over 500 tons of fish are caught, the actual tonnage will be over 400 as well. The event that the town's fleet will catch between 400 and 500 tons of fish is $F \setminus E$ since E hasn't occurred, but F has. So using property (4) from above we have $P(F \setminus E) = P(F) - P(E) = 0.35 - 0.1 = 0.25$. In other words there is a 25% chance that between 400 and 500 tons of fish will be caught. ■

Example 1.1.7. Suppose we know there is a 60% chance that it will rain tomorrow and a 70% chance the high temperature will be above 30^oC. Suppose we also know that there is a 40% chance that the high temperature will be above 30^oC and it will rain. How likely is it tomorrow will be a dry day that does not go above 30^oC?

The answer to this question may not be so obvious, but our first step is still to view the pieces of information in terms of events and probabilities. We have one event E which represents "It will rain tomorrow" and another F which represents "The high will be above 30^oC tomorrow". Our given probabilities tell us $P(E) = 0.6$, $P(F) = 0.7$, and $P(E \cap F) = 0.4$. We are trying to determine $P(E^c \cap F^c)$. We can do so using properties (5) and (6) proven above, together with the set-algebraic fact that $E^c \cap F^c = (E \cup F)^c$.

From (5) we know $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.7 + 0.6 - 0.4 = 0.9$. (This is the probability that it either will rain or be above 30 ^0C).

Then from (6) and the set-algebraic fact, $P(E^c \cap F^c) = P((E \cup F)^c) = 1 - P(E \cup F) = 1 - 0.9 = 0.1$.

So there is a 10% chance tomorrow will be a dry day that doesn't reach 30 degrees. ■

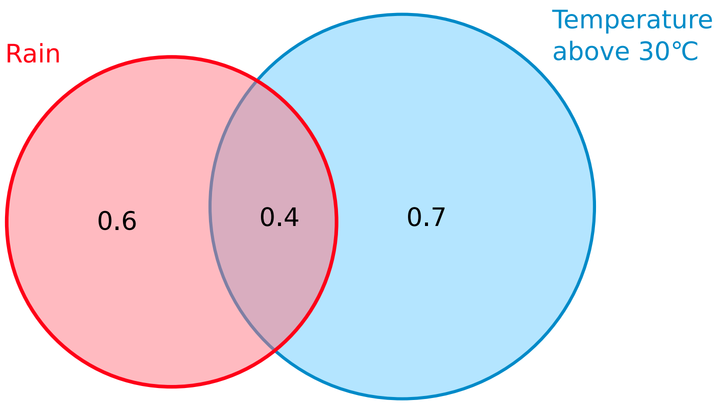

Figure 1.1: A Venn diagram that describes the probabilities from Example 1.1.7.

EXERCISES

Ex. 1.1.1. Consider the sample space $\Omega = \{a, b, c, d, e\}$. Given that $\{a, b, e\}$, and $\{b, c\}$ are both **events**, what other subsets of Ω must be events due to the requirement that the collection of events is closed under taking unions, intersections, and compliments?

Ex. 1.1.2. There are two positions - Cashier and Waiter - open at the local restaurant. There are two male applicants (David and Rajesh) two female applicants (Veronica and Megha). The Cashier position is chosen by selecting one of the four applicants at random. The Waiter position is then chosen by selecting at random one of the three remaining applicants.

(a) List the elements of the sample space S.

(b) List the elements of the event A that the position of Cashier is filled by a female applicant.

(c) List the elements of the event B that exactly one of the two positions is filled by a female applicant.

(d) List the elements of the event C that neither position was filled by a female applicant.

(e) Sketch a Venn diagram to show the relationship among the events A, B, C and S.

Ex. 1.1.3. A jar contains a large collection of red, green, and white marbles. Marbles are drawn from the jar one at a time. The color of the marble is recorded and it is put back in the jar before the next draw. Let R_n denote the event that the n^{th} draw is a red marble and let G_n denote the event that the n^{th} draw is a green marble. For example, $R_1 \cap G_2$ would denote the event that the first marble was red and the second was green. In terms of these events (and appropriate set-theoretic symbols – union, intersection, and complement) find expressions for the events in parts (a), (b), and (c) below.

(a) The first marble drawn is white. (We might call this W_1, but show that it can be written in terms of the R_n and G_n sets described above).

(b) The first marble drawn is green and the second marble drawn is not white.

(c) The first and second draws are different colors.

(d) Let $E = R_1 \cup G_2$ and let $F = R_1^c \cap R_2$. Are E and F disjoint? Why or why not?

Ex. 1.1.4. Suppose there are only thirteen teams with a non-zero chance of winning the next World Cup. Suppose those teams are Spain (with a 14% chance), the Netherlands (with a 11% chance), Germany (with a 11% chance), Italy (with a 10% chance), Brazil (with a 10% chance), England (with a 9% chance), Argentina (with a 9% chance), Russia (with a 7% chance), France (with an 6% chance), Turkey (with a 4% chance), Paraguay (with a 4% chance), Croatia (with a 4% chance) and Portugal (with a 1% chance).

(a) What is the probability that the next World Cup will be won by a South American country?

(b) What is the probability that the next World Cup will be won by a country that is not from South America? (Think of two ways to do this problem - one directly and one using part (5) of Theorem 1.1.4. Which do you prefer and why?)

Ex. 1.1.5. If A and B are disjoint events and $P(A) = 0.3$ and $P(B) = 0.6$, find $P(A \cup B)$, $P(A^c)$ and $P(A^c \cap B)$.

Ex. 1.1.6. Suppose E and F are events in a sample space S. Suppose that $P(E) = 0.7$ and $P(F) = 0.5$.

(a) What is the largest possible value of $P(E \cap F)$? Explain.

(b) What is the smallest possible value of $P(E \cap F)$? Explain.

Ex. 1.1.7. A biologist is modeling the size of a frog population in a series of ponds. She is concerned with both the number of egg masses laid by the frogs during breeding season and the annual precipitation into the ponds. She knows that in a given year there is an 86% chance that there will be over 150 egg masses deposited by the frogs (event E) and that there is a 64% chance that the annual precipitation will be over 17 inches (event F).

(a) In terms of E and F, what is the event "there will be over 150 egg masses and an annual precipitation of over 17 inches"?

(b) In terms of E and F, what is the event "there will be 150 or fewer egg masses and the annual precipitation will be over 17 inches"?

(c) Suppose the probability of the event from (a) is 59%. What is the probability of the event from (b)?

Ex. 1.1.8. In part (6) of Theorem 1.1.4 we showed that

$$P(E \cup F) = P(E) + P(F) - P(E \cap F).$$

Versions of this rule for three or more sets are explored below.

(a) Prove that $P(A \cup B \cup C)$ is equal to

$$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

for any events A, B, and C.

(b) Use part (a) to answer the following question. Suppose that in a certain United States city 49.3% of the population is male, 11.6% of the population is sixty-five years of age or older, and 13.8% of the population is Hispanic. Further, suppose 5.1% is both male and at least sixty-five, 1.8% is both male and Hispanic, and 5.9% is Hispanic and at least sixty-five. Finally, suppose that 0.7% of the population consists of Hispanic men that are at least sixty-five years old. What percentage of people in this city consists of non-Hispanic women younger than sixty-five years old?

(c) Find a four-set version of the equation. That is, write $P(A \cup B \cup C \cup D)$ in terms of probabilities of intersections of A, B, C, and D.

(d) Find an n-set version of the equation.

Ex. 1.1.9. A and B are two events. P(A)=0.4, P(B)=0.3, P(A∪B)=0.6. Find the following probabilities:

(a) $P(A \cap B)$;

(b) P(Only A happens); and

(c) P(Exactly one of A or B happens).

Ex. 1.1.10. In the next subsection we begin to look at probability spaces where each of the outcomes are equally likely. This problem will help develop some early intuition for such problems.

(a) Suppose we roll a die and so $S = \{1, 2, 3, 4, 5, 6\}$. Each outcome separately $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ is an event. Suppose each of these events is equally likely. What must the probability of each event be? What axioms or properties are you using to come to your conclusion?

(b) With the same assumptions as in part (a), how would you determine the probability of an event like $E = \{1, 3, 4, 6\}$? What axioms or properties are you using to come to your conclusion?

(c) If $S = \{1, 2, 3, ..., n\}$ and each single-outcome event is equally likely, what would be the probability of each of these events?

(d) Suppose $E \subset S$ is an event in the sample space from part (c). Explain how you could determine $P(E)$.

Ex. 1.1.11. Suppose A and B are subsets of a sample space Ω.

(a) Show that $(A - B) \cup B = A$ when $B \subset A$.

(b) Show by example that the equality doesn't always hold if B is not a subset of A.

Ex. 1.1.12. Let A and B be events.

(a) Suppose $P(A) = P(B) = 0$. Prove that $P(A \cup B) = 0$.

(b) Suppose $P(A) = P(B) = 1$. Prove that $P(A \cap B) = 1$.

Ex. 1.1.13. Let A_n be a sequence of events.

(a) Suppose $A_n \subseteq A_{n+1}$ for all $n \geq 1$. Show that

$$P(\cup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n)$$

(b) Suppose $A_n \supseteq A_{n+1}$ for all $n \geq 1$. Show that

$$P(\cap_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n)$$

1.2 EQUALLY LIKELY OUTCOMES

When a sample space S consists of only a countable collection of outcomes, describing the probability of each individual outcome is sufficient to describe the probability of all events. This is because if $A \subset S$ we may simply compute

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

This assignment of probabilities to each outcome is called a "distribution" since it describes how probability is distributed amongst the possibilities. Perhaps the simplest example arises when there are a finite collection of equally likely outcomes. Think of examples such as flipping a fair coin ("heads" and "tails" are equally likely to occur), rolling a fair die (1, 2, 3, 4, 5, and 6 are equally likely), or drawing a set of numbers for a lottery (many possibilities, but in a typical lottery, each outcome is as likely as any other). Such distributions are common enough that it is useful to have shorthand notations for them. In the case of a sample space $S = \{\omega_1, \ldots, \omega_n\}$ where each outcome is equally likely, the probability is referred to as a "uniform distribution" and is denoted by Uniform$(\{\omega_1, \ldots, \omega_n\})$. In such situations, computing probabilities simply reduces to computing the number of outcomes in a given event and consequently becomes a combinatorial problem.

THEOREM 1.2.1. **Uniform**$(\{\omega_1, \ldots, \omega_n\})$ *: Let $S = \{\omega_1, \omega_2, ..., \omega_n\}$ be a non-empty, finite set. If $E \subset S$ is any subset of S, let $P(E) = \frac{|E|}{|S|}$ (where $|E|$ represents the number of elements of E). Then P defines a probability on S and P assigns equal probability to each individual outcome in S.*

Proof - Since $E \subset S$ we know $|E| \le |S|$ and so $0 \le P(E) \le 1$, so we must prove that P satisfies the two probability axioms.
Since $P(S) = \frac{|S|}{|S|} = 1$ the first axiom is satisfied.
To verify the second axiom, suppose $E_1, E_2, ...$ is a countable collection of disjoint events. Since S is finite, only finitely many of these E_j can be non-empty, so we may list the non-empty events as $E_1, E_2, \ldots, E_n$. For $j > n$ we know $E_j = \varnothing$ and so $P(E_j) = 0$ by the definition. Since the events are disjoint, to find the number of elements in their union we simply add the elements of each event separately. That is, $|E_1 \cup E_2 \cup \cdots \cup E_n| = |E_1| + |E_2| + \cdots + |E_n|$ and therefore

$$P(\bigcup_{j=1}^{\infty} E_j) = \frac{|\cup_{j=1}^{\infty} E_j|}{|S|} = \frac{\sum_{j=1}^{n} |E_j|}{|S|} = \sum_{j=1}^{n} \frac{|E_j|}{|S|} = \sum_{j=1}^{n} P(E_j) = \sum_{j=1}^{\infty} P(E_j).$$

Finally, let $\omega \in S$ be any single outcome and let $E = \{\omega\}$. Then $P(E) = \frac{1}{|S|}$, so every outcome in S is equally likely. ■

EXAMPLE 1.2.2. A deck of twenty cards labeled 1, 2, 3, ..., 20 is shuffled and a card selected at random. What is the probability that the number on the card is a multiple of six?

The description of the scenario suggests that each of the twenty cards is as likely to be chosen as any other. In this case $S = \{1, 2, 3, ..., 20\}$ while $E = \{6, 12, 18\}$. Therefore, $P(E) = \frac{|E|}{|S|} = \frac{3}{20} = 0.15$. There is a 15% chance that the card will be a multiple of six. ■

EXAMPLE 1.2.3. Two dice are rolled. How likely is it that their sum will equal eight?

Since we are looking at a sum of dice, it might be tempting to regard the sample space as $S = \{2, 3, 4, ..., 11, 12\}$, the collection of possible sums. While this is a possible approach (and one we will return to later), it is not the case that all of these outcomes are equally likely. Instead we can view an experiment as tossing a first die and a second die and recording the pair of numbers that occur on each of the dice. Each of these pairs is as likely as any other to occur. So

$$S = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

and $|S| = 6 \times 6 = 36$. The event that the sum of the dice is an eight is $E = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$. Therefore $P(E) = \frac{|E|}{|S|} = \frac{5}{36}$. ■

Example 1.2.4. A seven letter code is selected at random with every code as likely to be selected as any other code (so AQRVTAS and CRXAOLZ would be two possibilities). How likely is it that such a code has at least one letter used more than once? (This would happen with the first code above with a repeated A - but not with the second).

As with the examples above, the solution amounts to counting numbers of outcomes. However, unlike the examples above the numbers involved here are quite large and we will need to use some combinatorics to find the solution. The sample space S consists of all seven-letter codes from AAAAAAA to ZZZZZZZ. Each of the seven spots in the code could be any of twenty-six letters, so $|S| = 26^7 = 8,031,810,176$. If E is the event for which there is at least one letter used more than once, it is easier to count E^c, the event where no letter is repeated. Since in this case each new letter rules out a possibility for the next letter there are $26 \times 25 \times 24 \times 23 \times 22 \times 21 \times 20 = 3,315,312,000$ such possibilities.

This lets us compute $P(E^c) = \frac{3{,}315{,}312{,}000}{8{,}031{,}810{,}176}$ from which we find $P(E) = 1 - P(E^c) = \frac{4{,}716{,}498{,}176}{8{,}031{,}810{,}176} \approx 0.587$. That is, there is about a 58.7% chance that such a code will have a repeated letter. ■

Example 1.2.5. A group of twelve people includes Grant and Dilip. A group of three people is to be randomly selected from the twelve. How likely is it that this three-person group will include Grant, but not Dilip?

Here, S is the collection of all three-person groups, each of which is as likely to be selected as any other. The number of ways of selecting a three-person group from a pool of twelve is $|S| = \binom{12}{3} = 220$. The event E consists of those three-person groups that include Grant, but not Dilip. Such groups must include two people other than Grant and there are ten people remaining from which to select the two, so $|E| = \binom{10}{2} = 45$. Therefore, $P(E) = \frac{45}{220} = \frac{9}{44}$. ■

EXERCISES

Ex. 1.2.1. A day is selected at random from a given week with each day as likely to be selected as any other.

(a) What is the sample space S? What is the size of S?

(b) Let E be the event that the selected day is a Saturday or a Sunday. What is the probability of E.

Ex. 1.2.2. A box contains 500 envelopes, of which 50 contain Rs 100 in cash, 100 contain Rs 50 in cash and 350 contain Rs 10. An envelope can be purchased at Rs 25 from the owner, who will pick an envelope at random and give it to you. Write down the sample space for the net money gained by you. If each envelope is as likely to be selected as any other envelope, what is the probability that the first envelope purchased contains less than Rs 100?

Ex. 1.2.3. Three dice are tossed.

(a) Describe (in words) the sample space S and give an example of an object in S.

(b) What is the size of S?

(c) Let E be the event that the first two dice both come up "1". What is the size of E? What is the probability of E?

(d) Let G be the event that the three dice show three different numbers. What is the size of G? What is the probability of G?

(e) Let F be the event that the third die is larger than the sum of the first two. What is the size of F? What is the probability of F?

Ex. 1.2.4. Suppose that each of three women at a party throws her hat into the center of the room. The hats are first mixed up and then each one randomly selects a hat. Describe the probability space for the possible selection of hats. If all of these selections are equally likely, what is the probability that none of the three women selects her own hat?

Ex. 1.2.5. A group of ten people includes Sona and Adam. A group of five people is to be randomly selected from the ten. How likely is it that this group of five people will include neither Sona nor Adam?

Ex. 1.2.6. There are eight students with two females and six males. They are split into two groups A and B, of four each.

(a) In how many different ways can this be done?

(b) What is the probability that two females end up in group A?

(c) What is the probability that there is one female in each group?

Ex. 1.2.7. Sheela has lost her key to her room. The security officer gives her 50 keys and tells her that one of them will open her room. She decides to try each key successively and notes down the number of the attempt at which the room opens. Describe the sample space for this experiment. Do you think it is realistic that each of these outcomes is equally likely? Why or why not?

Ex. 1.2.8. Suppose that n balls, of which k are red, are arranged at random in a line. What is the probability that all the red balls are next to each other?

Ex. 1.2.9. Consider a deck of 50 cards. Each card has one of 5 colors (black, blue, green, red, and yellow), and is printed with a number (1,2,3,4,5,6,7,8,9, or 10) so that each of the 50 color/number combinations is represented exactly once. A **hand** is produced by dealing out five different cards from the deck. The order in which the cards were dealt does not matter.

(a) How many different hands are there?

(b) How many hands consist of cards of identical color? What is the probability of being dealt such a hand?

(c) What is the probability of being dealt a hand that contains exactly three cards with one number, and two cards with a different number?

(d) What is the probability of being dealt a hand that contains two cards with one number, two cards with a different number, and one card of a third number?

Ex. 1.2.10. Suppose you are in charge of quality control for a light bulb manufacturing company. Suppose that in the process of producing 100 light bulbs, either all 100 bulbs will work properly, or through some manufacturing error twenty of the 100 will not work. Suppose your quality control procedure is to randomly select ten bulbs from a 100 bulb batch and test them to see if they work properly. How likely is this procedure to detect if a batch has bad bulbs in it?

Ex. 1.2.11. A fair die is rolled five times. What is the probability of getting at least two 5's and at least two 6's among the five rolls.

Ex. 1.2.12. **(The "Birthday Problem")** For a group of N people, if their birthdays were listed one-by-one, there are 365^N different ways that such a list might read (if we ignore February 29 as a possibility). Suppose each of those possible lists is as likely as any other.

(a) For a group of two people, let E be the event that they have the same birthday. What is the size of E? What is the probability of E?

(b) For a group of three people, let F be the event that at least two of the three have the same birthday. What is the size of F? What is the probability of F? (Hint: It is easier to find the size of F^c than it is to find the size of F).

(c) For a group of four people, how likely is it that at least two of the four have the same birthday?

(d) How large a group of people would you need to have before it becomes more likely than not that at least two of them share a birthday?

Ex. 1.2.13. A coin is tossed 100 times.

(a) How likely is it that the 100 tosses will produce exactly fifty heads and fifty tails?

(b) How likely is it that the number of heads will be between 50 and 55 (inclusive)?

Ex. 1.2.14. Suppose I have a coin that I claim is "fair" (equally likely to come up heads or tails) and that my friend claims is weighted towards heads. Suppose I flip the coin twenty times and find that it comes up heads on sixteen of those twenty flips. While this seems to favor my friend's hypothesis, it is still possible that I am correct about the coin and that just by chance the coin happened to come up heads more often than tails on this series of flips. Let S be the sample space of all possible sequences of flips. The size of S is then 2^{20}, and if I am correct about the coin being "fair", each of these outcomes are equally likely.

(a) Let E be the event that exactly sixteen of the flips come up heads. What is the size of E? What is the probability of E?

(b) Let F be the event that at least sixteen of the flips come up heads. What is the size of F? What is the probability of F?

Note that the probability of F is the chance of getting a result as extreme as the one I observed if I happen to be correct about the coin being fair. The larger $P(F)$ is, the more reasonable seems my assumption about the coin being fair. The smaller $P(F)$ is, the more that assumption looks doubtful. This is the basic idea behind the statistical concept of "hypothesis testing" which we will revisit in Chapter 9.

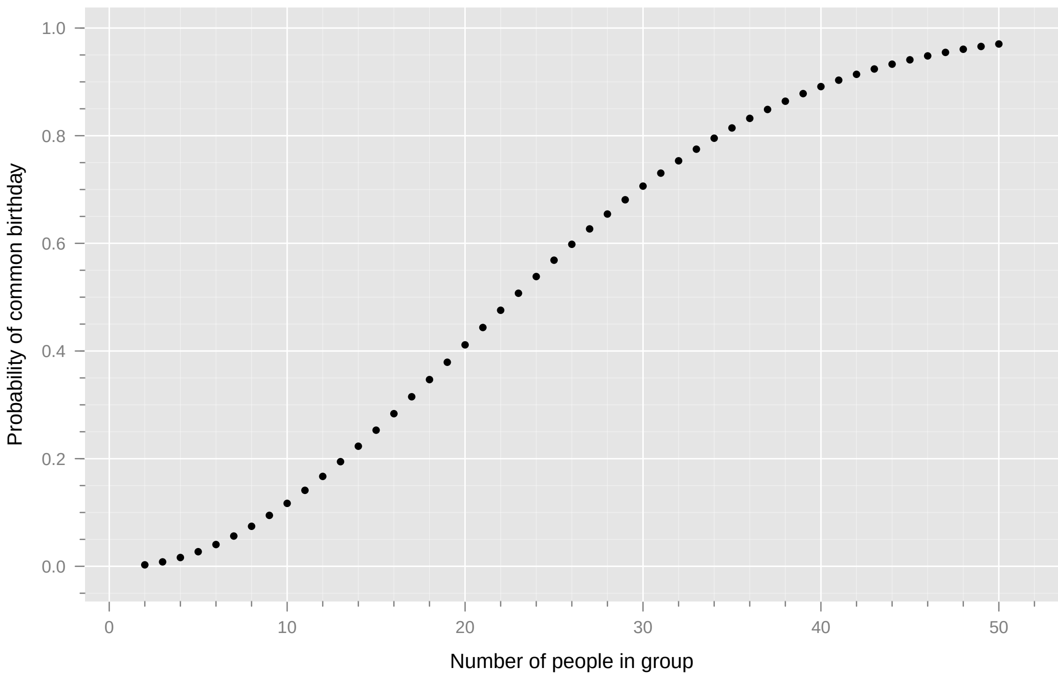

Figure 1.2: The birthday problem discussed in Exercise 1.2.12

Ex. 1.2.15. Suppose that r indistinguishable balls are placed in n distinguishable boxes so that each distinguishable arrangement is equally likely. Find the probability that no box will be empty.

Ex. 1.2.16. Suppose that 10 potato sticks are broken into two - one long and one short piece. The 20 pieces are now arranged into 10 random pairs chosen uniformly.

(a) Find the probability that each of pairs consists of two pieces that were originally part of the same potato stick.

(b) Find the probability that each pair consists of a long piece and a short piece.

Ex. 1.2.17. Let S be a non-empty, countable (finite or infinite) set such that for each $\omega \in S, 0 \leq p_\omega \leq 1$. Let $\mathcal{F}$ be the collection of all events. Suppose $P : \mathcal{F} \to [0, 1]$ is given by

$$P(E) = \sum_{\omega \in E} p_\omega,$$

for any event E.

(a) Show that P satisfies Axiom 2 in Definition 1.1.3.

(b) Further, conclude that if $P(S) = 1$ then P defines a probability on S.

1.3 CONDITIONAL PROBABILITY AND BAYES' THEOREM

In the previous section we introduced an axiomatic definition of "probability" and discussed the concept of an "event". Now we look at ways in which the knowledge that one event has occurred may be used as information to inform and alter the probability of another event.

EXAMPLE 1.3.1. Consider the experiment of tossing a fair coin three times with sample space $S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$. Let A be the event that there are two or more heads. As all outcomes are equally likely,

$$\begin{aligned} P(A) &= \frac{|A|}{|S|} \\ &= \frac{|\{hhh, hht, hth, thh\}|}{8} \\ &= \frac{1}{2}. \end{aligned}$$

Let B be the event that there is a head in the first toss. As above,

$$\begin{aligned} P(B) &= \frac{|B|}{|S|} \\ &= \frac{|\{hhh, hht, hth, htt\}|}{8} \\ &= \frac{1}{2}. \end{aligned}$$

Now suppose we are asked to find the probability of at least two or more heads among the three tosses, but we are also given the additional information that the first toss was a head. In other words, we are asked to find the probability of A, given the information that event B has definitely occurred. Since the additional information guarantees B is now a list of all possible outcomes, it makes intuitive sense to view the event B as a new sample space and then identify the subset $A \cap B = \{hhh, hht, hth\}$ of B consisting of outcomes for which there are at least two heads. We could conclude that the probability of at least two or more heads in three tosses given that the first toss was a head is

$$\frac{|A \cap B|}{|B|} = \frac{3}{4}.$$

■

This is a legitimate way to view the problem and it leads to the correct solution. However, this method has one very serious drawback – it requires us to change both our sample space and our probability function in order to carry out the computation. It would be preferable to have a method that allows us to work within the original framework of the sample space S and to talk about the "conditional probability" of A given that the result of the experiment will be an outcome in B. This is denoted as $P(A|B)$ and is read as "the (conditional) probability of A given B."

Suppose S is a finite set of equally likely outcomes from a given experiment. Then for any two non-empty events A and B, the conditional probability of A given B is given by

$$\frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|S|}}{\frac{|B|}{|S|}} = \frac{P(A \cap B)}{P(B)}.$$

This leads us to a formal definition of conditional probability for general sample spaces.

DEFINITION 1.3.2. **(Conditional Probability)** *Let S be a sample space with probability P. Let A and B be two events with $P(B) > 0$. Then the conditional probability of A given B written as $P(A|B)$ and is defined by*

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

This definition makes it possible to compute a conditional probability in terms of the original (unconditioned) probability function.

Example 1.3.3. A pair of dice are thrown. If it is known that one die shows a 4, what is the probability the other die shows a 6?

Let B be the event that one of the dice shows a 4. So

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

Let A be the event that one of the dice is a 6. So

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

then

$$A \cap B = \{(4,6),(6,4)\}.$$

Hence

$$\begin{aligned}
P(\text{one die shows 6} \quad &\Big| \quad \text{one die shows 4}) \\
&= P(A|B) = \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(\text{one die shows 6 and the other shows 4})}{P(\text{one die shows 4})} \\
&= \frac{2/36}{11/36} = \frac{2}{11}.
\end{aligned}$$

■

In many applications, the conditional probabilities are implicitly defined within the context of the problem. In such cases, it is useful to have a method for computing non-conditional probabilities from the given conditional ones. Two such methods are given by the next results and the subsequent examples.

Example 1.3.4. An economic model predicts that if interest rates rise, then there is a 60% chance that unemployment will increase, but that if interest rates do not rise, then there is only a 30% chance that unemployment will increase. If the economist believes there is a 40% chance that interest rates will rise, what should she calculate is the probability that unemployment will increase?

Let B be the event that interest rates rise and A be the event that unemployment increases. We know the values

$$P(B) = 0.4,\ P(B^c) = 0.6,\ P(A|B) = 0.6,\ \text{and}\ P(A|B^c) = 0.3.$$

Using the axioms of probability and definition of conditional probability we have

$$\begin{aligned}
P(A) &= P((A \cap B) \cup (A \cap B^c)) \\
&= P((A \cap B)) + P(A \cap B^c)) \\
&= P(A|B)P(B) + P(A|B^c)P(B^c) \\
&= 0.6(0.4) + 0.3(0.6) = 0.42.
\end{aligned}$$

So there is a 42% chance that unemployment will increase. ■

THEOREM 1.3.5. *Let A be an event and let $\{B_i : 1 \le i \le n\}$ be a disjoint collection of events for which $P(B_i) > 0$ for all i and such that $A \subset \bigcup_{i=1}^{n} B_i$. Suppose $P(B_i)$ and $P(A|B_i)$ are known. Then $P(A)$ may be computed as*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Proof - The events $(A \cap B_i)$ and $(A \cap B_j)$ are disjoint if $i \neq j$ and

$$\bigcup_{i=1}^{n}(A \cap B_i) = A \cap (\bigcup_{i=1}^{n} B_i) = A.$$

So,

$$\begin{aligned} P(A) &= P(\bigcup_{i=1}^{n}(A \cap B_i)) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i). \end{aligned}$$

■

A nearly identical proof holds when there are only countably many B_i (see Exercise 1.3.11).

EXAMPLE 1.3.6. Suppose we have coloured balls distributed in three boxes in quantities as given by the table below:

	Box 1	Box 2	Box 3
Red	4	3	3
Green	3	3	4
Blue	5	2	3

A box is selected at random. From that box a ball is selected at random. How likely is it that a red ball is drawn?

Let B_1, B_2, and B_3 be the events that Box 1, 2, or 3 is selected, respectively. Note that these events are disjoint and cover all possibilities in the sample space. Let R be the event that the selected ball is red. Then by Theorem 1.3.5,

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\ &= \frac{4}{12} \cdot \frac{1}{3} + \frac{3}{8} \cdot \frac{1}{3} + \frac{3}{10} \cdot \frac{1}{3} \\ &= \frac{121}{360}. \end{aligned}$$

■

EXAMPLE 1.3.7. **(Polya's Urn Scheme)** Suppose there is an urn that contains r red balls and b black balls. A ball is drawn at random and its colour noted. It is replaced with $c > 0$ balls of the same colour. The procedure is then repeated. For $j = 1, 2, \ldots$, let R_j and B_j be the events that the j-th ball drawn is red and black respectively. Clearly $P(R_1) = \frac{r}{b+r}$ and $P(B_1) = \frac{b}{b+r}$. When the first ball is replaced, c new balls will be added to the urn, so that when the second ball is drawn there will be $r+b+c$ balls available. From this it can easily be checked that $P(R_2|R_1) = \frac{r+c}{b+r+c}$ and $P(R_2|B_1) = \frac{r}{b+r+c}$. Noting that R_1 and B_1 are disjoint and together represent the entire sample space, $P(R_2)$ can be computed as

$$\begin{aligned}
P(R_2) &= P(R_1)P(R_2|R_1) + P(B_1)P(R_2|B_1) \\
&= \frac{r}{b+r} \cdot \frac{r+c}{b+r+c} + \frac{b}{b+r} \cdot \frac{r}{b+r+c} \\
&= \frac{r(r+b+c)}{(r+b+c)(b+r)} \\
&= \frac{r}{b+r} = P(R_1).
\end{aligned}$$

One can show that $P(R_j) = \frac{r}{b+r}$ for all $j \geq 1$.

The urn schemes were originally developed by George Polya (1887-1985). Various modifications to Polya's urn scheme are discussed in the exercises. ■

Above we have described how conditioning on an event B may be viewed as modifying the original probability based on the additional information provided by knowing that B has occurred. Frequently in applications we gain information more than once in the process of an experiment. The following theorem shows how to deal with such a situation.

Theorem 1.3.8. *For an integer $n \geq 2$, let $A_1, A_2, \dots, A_n$ be a collection of events for which $\bigcap_{j=1}^{n-1} A_j$ has positive probability. Then,*

$$P(\bigcap_{j=1}^{n} A_j) = P(A_1) \cdot \prod_{j=2}^{n} P(A_j \,|\, \bigcap_{k=1}^{j-1} A_k).$$

The proof of this theorem is left as Exercise 1.3.14, but we will provide a framework in which to make sense of the equality. Usually the events $A_1, \dots, A_n$ are viewed as a sequence in time for which we know the probability of a given event provided that all of the others before it have already occurred. Then we can calculate $P(A_1 \cap A_2 \cap \cdots \cap A_n)$ by taking the product of the values $P(A_1), P(A_2|A_1), P(A_3|A_1 \cap A_2), \dots, P(A_n|A_1 \cap \cdots \cap A_{n-1})$.

Example 1.3.9. A probability class has fifteen students - four seniors, eight juniors, and three sophomores. Three different students are selected at random to present homework problems. What is the probability the selection will be a junior, a sophomore, and a junior again, in that order?

Let A_1 be the event that the first selection is a junior. Let A_2 be the event that the second selection is a sophomore, and let A_3 be the event that the third selection is a junior. The problem asks for $P(A_1 \cap A_2 \cap A_3)$ which we can calculate using Theorem 1.3.8.

$$\begin{aligned}
P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\
&= \frac{8}{15} \cdot \frac{3}{14} \cdot \frac{7}{13} \\
&= \frac{4}{65}.
\end{aligned}$$

■

1.3.1 *Bayes' Theorem*

It is often the case that we know the conditional probability of A given B, but want to know the conditional probability of B given A instead. It is possible to calculate one quantity from the other using a formula known as Bayes' theorem. We introduce this with a motivating example.

EXAMPLE 1.3.10. We return to Example 1.3.6. In that example we had three boxes containing balls given by the table below:

	Box 1	Box 2	Box 3
Red	4	3	3
Green	3	3	4
Blue	5	2	3

A box is selected at random. From the box a ball is selected at random. When we looked at conditional probabilities we saw how to determine the probability of an event such as {the ball drawn is red}. Now suppose we know the ball is red and want to determine the probability of the event {the ball was drawn from box 3}. That is, if R is the event that a red ball is chosen and if B_1, B_2, and B_3 are the events that boxes 1, 2, and 3 are selected, we want to determine the conditional probability $P(B_3|R)$. The difficulty is that while the conditional probabilities $P(R|B_1), P(R|B_2)$, and $P(R|B_3)$ are easy to determine, calculating the conditional probability with the order of the events reversed is not immediately obvious.

Using the definition of conditional probability we have that

$$P(B_3|R) = \frac{P(B_3 \cap R)}{P(R)}.$$

We can rewrite

$$P(B_3 \cap R) = P(R|B_3)P(B_3) = (\frac{3}{10})(\frac{1}{3}) = 0.1.$$

On the other hand, we can decompose the event R over which box was chosen. This is exactly what we did to solve Example 1.3.6 where we found that $P(R) = \frac{121}{360}$. Hence,

$$P(B_3|R) = \frac{P(B_3 \cap R)}{P(R)} = \frac{0.1}{121/360} = \frac{36}{121} \approx 0.298.$$

So if we know that a red ball was drawn, there is slightly less than a 30% chance that it came from Box 3. ■

In the above example the description of the experiment allowed us to determine $P(B_1)$, $P(B_2)$, $P(B_3)$, $P(R|B_1)$, $P(R|B_2)$, and $P(R|B_3)$. We were then able to use the definition of conditional probability to find $P(B_3|R)$. Such a computation can be done in general.

THEOREM 1.3.11. **(Bayes' Theorem)** *Suppose A is an event, $\{B_i : 1 \leq i \leq n\}$ are a collection of disjoint events whose union contains all of A. Further assume that $P(A) > 0$ and $P(B_i) > 0$ for all $1 \leq i \leq n$. Then for any $1 \leq i \leq n$,*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Proof -

$$\begin{aligned} P(B_i \mid A) &= \frac{P(B_i \cap A)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{P(\bigcup_{j=1}^{n} A \cap B_j)} \\ &= \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \cap B_j)} \\ &= \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}. \end{aligned} \tag{1.3.1}$$

■

The equation (1.3.1) is sometimes referred to as "Bayes' formula" or "Bayes' rule" as well. This is originally due to Thomas Bayes (1701-1761).

EXAMPLE 1.3.12. Shyam is randomly selected from the citizens of Hyderabad by the Health authorities. A laboratory test on his blood sample tells Shyam that he has tested positive for Swine Flu. It is found that 95% of people with Swine Flu test positive but 2% of people without the disease will also test positive. Suppose that 1% of the population has the disease. What is the probability that Shyam indeed has the Swine Flu ? ■

Consider the events $A = \{$ Shyam has Swine Flu $\}$ and $B = \{$ Shyam tested postive for Swine Flu $\}$. We are given:

$$P(B|A) = 0.95,\ P(B|A^c) = 0.02,\ \text{ and } P(A) = 0.01.$$

Using Bayes' Theorem we have,

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{(0.95)(0.01)}{(0.95)(0.01) + (0.02)(0.99)} \\ &= 0.324 \end{aligned}$$

Despite testing positive, there is only a 32.4 percent chance that Shyam has the disease.

EXERCISES

Ex. 1.3.1. There are two dice, one red and one blue, sitting on a table. The red die is a standard die with six sides while the blue die is tetrahedral with four sides, so the outcomes 1, 2, 3, and 4 are all equally likely. A fair coin is flipped. If that coin comes up heads, the red die will be rolled, but if the coin comes up tails the blue die will be rolled.

(a) Find the probability that the rolled die will show a 1.

(b) Find the probability that the rolled die will show a 6.

Ex. 1.3.2. A pair of dice are thrown. It is given that the outcome on one die is a 3. what is the probability that the sum of the outcomes on both dice is greater than 7?

Ex. 1.3.3. Box A contains four white balls and three black balls and Box B contains three white balls and five black balls.

(a) Suppose a box is selected at random and then a ball is chosen from the box. If the ball drawn is black then what is the probability that it was from Box A?

(b) Suppose instead that one ball is drawn at random from Box A and placed (unseen) in Box B. What is the probability that a ball now drawn from Box B is black?

Ex. 1.3.4. Tomorrow the weather will either be sunny, cloudy, or rainy. There is a 60% chance tomorrow will be cloudy, a 30% chance tomorrow will be sunny, and a 10% chance that tomorrow will be rainy. If it rains, I will not go on a walk. But if it is cloudy, there is a 90% chance I will take a walk and if it's sunny there is a 70% chance I will take a walk. If I take a walk on a cloudy day, there is an 80% chance I will walk further than five kilometers, but if I walk on a sunny day, there's only a 50% chance I will walk further than five kilometers. Using the percentages as given probabilities, answer the following questions:

(a) How likely is it that tomorrow will be cloudy and I will walk over five kilometers?

(b) How likely is it I will take a walk over five kilometers tomorrow?

Ex. 1.3.5. A box contains B black balls and W balls, where $W \geq 3, B \geq 3$. A sample of three balls is drawn at random with each drawn ball being discarded (not put back into the box) after it is drawn. For $j = 1, 2, 3$ let A_j denote the event that the ball drawn on the j^{th} draw is white. Find $P(A_1)$, $P(A_2)$ and $P(A_3)$.

Ex. 1.3.6. There are two sets of cards, one red and one blue. The red set has four cards - one that reads 1, two that read 2, and one that reads 3. An experiment involves flipping a fair coin. If the coin comes up heads a card will be randomly selected from the red set (and its number recorded) while if the coin comes up tails a card will be randomly selected from the blue set (and its number recorded). You can construct the blue set of cards in any way you see fit using any number of cards reading 1, 2, or 3. Explain how to build the blue set of cards to make each of the experimental outcomes $1, 2, 3$ equally likely.

Ex. 1.3.7. There are three tables, each with two drawers. Table 1 has a red ball in each drawer. Table 2 has a blue ball in each drawer. Table 3 has a red ball in one drawer and a blue ball in the other. A table is chosen at random, then a drawer is chosen at random from that table. Find the conditional probability that Table 1 is chosen, given that a red ball is drawn.

Ex. 1.3.8. In the G.R.E advanced mathematics exam, each multiple choice question has 4 choices for an answer. A prospective graduate student taking the test knows the correct answer with probability $\frac{3}{4}$. If the student does not know the answer, she guesses randomly. Given that a question was answered correctly, find the conditional probability that the student knew the answer.

Ex. 1.3.9. You first roll a fair die, then toss as many fair coins as the number that showed on the die. Given that 5 heads are obtained, what is the probability that the die showed 5 ?

Ex. 1.3.10. Manish is a student in a probability class. He gets a note saying, "I've organized a probability study group tonight at 7pm in the coffee shop. Come if you want." The note is signed "Hannah". However, Manish has class with two different Hannahs and he isn't sure which one sent the note. He figures that there is a 75% chance that Hannah A. would have organized such a study group, but only a 25% chance that Hannah B. would have done so. However, he also figures that if Hannah A. had organized the group, there is an 80% chance that she would have planned to meet on campus and only a 20% chance that she would have planned to meet in the coffee shop. While if Hannah B. had organized the group there is a 10% chance she would have planned for it on campus and a 90% chance she would have chosen the coffee shop. Given all this information, determine whether it is more likely that Manish should think the note came from Hannah A. or from Hannah B.

Ex. 1.3.11. State and prove a version of

(a) Theorem 1.3.5 when $\{B_i\}$ is a countably infinite collection of disjoint events.

(b) Theorem 1.3.11 when $\{B_i\}$ is a countably infinite collection of disjoint events.

Ex. 1.3.12. A bag contains 100 coins. Sixty of the coins are fair. The rest are biased to land heads with probability p (where $0 \le p \le 1$). A coin is drawn at random from the bag and tossed.

(a) Given that the outcome was a head what is the conditional probability that it is a biased coin?

(b) Evaluate your answer to (a) when $p = 0$. Can you explain why this answer should be intuitively obvious?

(c) Evaluate your answer to (a) when $p = \frac{1}{2}$. Can you explain why this answer should be fairly intuitive as well?

(d) View your answer to part (a) as a function $f(p)$. Show that $f(p)$ is an increasing function when $0 \le p \le 1$. Give an interpretation of this fact in the context of the problem.

Ex. 1.3.13. An urn contains b black balls and r red balls. A ball is drawn at random. The ball is replaced into the urn along with c balls of its colour and d balls of the opposite colour. Then another random ball is drawn and the procedure is repeated.

(a) What is the probability that the second ball drawn is a red ball?

(b) Assume $c = d$. What is the probability that the second ball drawn is a black ball?

(c) Still assuming $c = d$, what is the probability that the n^{th} ball drawn is a black ball?

(d) Assume $c > 0$ and $d = 0$, what is the probability that the n^{th} ball drawn is a black ball?

(6) Can you comment on the answers to (b) and/or (c) if the assumption that $c = d$ was removed?

Ex. 1.3.14. Use the following steps to prove Theorem 1.3.8.

(a) Prove Theorem 1.3.8 for the $n = 2$ case. (Hint: The proof should follow immediately from the definition of conditional probability).

(b) Prove Theorem 1.3.8 for the $n = 3$ case. (Hint: Rewrite the conditional probabilities in terms of ordinary probabilities).

(c) Prove Theorem 1.3.8 generally. (Hint: One method is to use induction, and parts (a) and (b) have already provided a starting point).

1.4 INDEPENDENCE

In the previous section we have seen instances where the probability of an event may change given the occurrence of a related event. However it is instructive and useful to study the case of two events where the occurrence of one has no effect on the probability of the other. Such events are said to be "independent".

EXAMPLE 1.4.1. Suppose we toss a coin three times. Then the sample space

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}.$$

Define $A = \{hhh, hht, hth, htt\} = \{$ the first toss is a head $\}$ and similarly define $B = \{hhh, hht, thh, tht\} = \{$ the second toss is a head $\}$. Note that $P(A) = \frac{1}{2} = P(B)$, while

$$P(A|B) = \frac{P(A \cap B)}{P(A)} = \frac{|A \cap B|}{|B|} = \frac{2}{4} = \frac{1}{2}$$

and

$$P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{|A \cap B^c|}{|B^c|} = \frac{2}{4} = \frac{1}{2}.$$

We have shown that $P(A) = P(A|B) = P(A|B^c)$. Therefore we conclude that the occurrence (or non-occurrence) of B has no effect on the probability of A. ■

This is the sort of condition we would want in a definition of independence. However, since defining $P(A|B)$ requires that $P(B) > 0$, our formal definition of "independence" will appear slightly different.

Definition 1.4.2. **(Independence)** *Two events A and B are independent if*

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

Example 1.4.3. Suppose we roll a die twice and denote as an ordered pair the result of the rolls. Suppose

$$E = \{ \text{ a six appears on the first roll } \} = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

and

$$F = \{ \text{ a six appears on the second roll } \} = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}.$$

As $E \cap F = \{(6,6)\}$, it is easy to see that

$$P(E \cap F) = \frac{1}{36}, P(E) = \frac{6}{36} = \frac{1}{6}, P(F) = \frac{6}{36} = \frac{1}{6}.$$

So E, F are independent as $P(E \cap F) = P(E)P(F)$. ■

Using the definition of conditional probability it is not hard to show (see Exercise 1.4.9) that if A and B are independent, and if $0 < P(B) < 1$ then

$$P(A|B) = P(A) = P(A|B^c). \tag{1.4.1}$$

If $P(A) > 0$ then the equations of (1.4.1) also hold with the roles of A and B reversed. Thus, independence implies four conditional probability equalities.

If we want to extend our definition of independence to three events A_1, A_2, and A_3, we would certainly want

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \tag{1.4.2}$$

to hold. We would also want any pair of the three events to be independent of each other. It is tempting to hope that pairwise independence is enough to imply (1.4.2). However, consider the following example.

Example 1.4.4. Suppose we toss a fair coin two times. Consider the three events $A_1 = \{hh, tt\}$, $A_2 = \{hh, ht\}$, and $A_3 = \{hh, th\}$. Then it is easy to calculate that

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2},$$

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}, \quad \text{and}$$

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4}.$$

So even though A_1, A_2 and A_3 are pairwise independent this does not imply that they satisfy (1.4.2). ■

It may also be tempting to hope that (1.4.2) is enough to imply pairwise independence, but that is not true either (see Exercise 1.4.6). The root of the problem is that, unlike the two event case, (1.4.2) does not imply that equality holds if any of the A_i are replaced by their complements. One solution is to insist that the multiplicative equality hold for any intersection of the events or their complements, which gives us the following definition.

Definition 1.4.5. **(Mutual Independence)** *A finite collection of events $A_1, A_2, \ldots, A_n$ is mutually independent if*

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)P(E_2)\ldots P(E_n). \tag{1.4.3}$$

whenever E_j is either A_j or A_j^c.

An arbitrary collection of events A_t where $t \in I$ for some index set I is mutually independent if every finite subcollection is mutually independent.

Thus, mutual independence of n events is defined in terms of 2^n equations. It is a fact (see Exercise 1.4.10) that if a collection of events is mutually independent, then so is any subcollection.

EXERCISES

Ex. 1.4.1. In the first semifinal of an international volleyball tournament Brazil has a 60% chance to beat Pakistan. In the other semifinal Poland has a 70% chance to beat Mexico. If the results of the two matches are independent, what is the probability that Pakistan will meet Poland in the tournament final?

Ex. 1.4.2. A manufacturer produces nuts and markets them as having 50mm radius. The machines that produce the nuts are not perfect. From repeated testing, it was established that 15% of the nuts have radius below 49mm and 12% have radius above 51mm. If two nuts are randomly (and independently) selected, find the probabilities of the following events:

(a) The radii of both the nuts are between 49mm and 51mm;

(b) The radius of at least one nut exceeds 51mm.

Ex. 1.4.3. Four tennis players (Avinash, Ben, Carlos, and David) play a single-elimination tournament with Avinash playing David and Ben playing Carlos in the first round and the winner of each of those contests playing each other in the tournament final. Below is the chart giving the percentage chance that one player will beat the other if they play. For instance, Avinash has a 30% chance of beating Ben if they happen to play.

	Avinash	Ben	Carlos	David
Avinash	-	30%	55%	40%
Ben	-	-	80%	45%
Carlos	-	-	-	15%
David	-	-	-	-

Suppose the outcomes of the games are independent. For each of the four players, determine the probability that player wins the tournament. Verify that the calculated probabilities sum to 1.

Ex. 1.4.4. Let A and B be events with $P(A) = 0.8$ and $P(B) = 0.7$.

(a) What is the largest possible value of $P(A \cap B)$?

(b) What is the smallest possible value of $P(A \cap B)$?

(c) What is the value of $P(A \cap B)$ if A and B are independent?

Ex. 1.4.5. Suppose we toss two fair dice. Let E_1 denote the event that the sum of the dice is six. E_2 denote the event that sum of the dice equals seven. Let F denote the event that the first die equals four. Is E_1 independent of F? Is E_2 independent of F?

Ex. 1.4.6. Suppose a bowl has twenty-seven balls. One ball is black, two are white, and eight each are green, red, and blue. A single ball is drawn from the bowl and its color is recorded. Define

$$\begin{aligned} A &= \{\text{the ball is either black or green}\} \\ B &= \{\text{the ball is either black or red}\} \\ C &= \{\text{the ball is either black or blue}\} \end{aligned}$$

(a) Calculate $P(A \cap B \cap C)$.

(b) Calculate $P(A)P(B)P(C)$.

(c) Are A, B, and C mutually independent? Why or why not?

Ex. 1.4.7. There are 150 students in the Probability 101 class. Of them, ninety are female, sixty use a pencil (instead of a pen), and thirty are wearing eye glasses. A student is chosen at random from the class. Define the following events:

$$\begin{aligned} A_1 &= \{\text{the student is a female}\} \\ A_2 &= \{\text{the student uses a pencil}\} \\ A_3 &= \{\text{the student is wearing eye glasses}\} \end{aligned}$$

(a) Show that it is impossible for these events to be mutually independent.

(b) Give an example to show that it may be possible for these events to be pairwise independent.

Ex. 1.4.8. When can an event be independent of itself? Do parts (a) and (b) below to answer this question.

(a) Prove that if an event A is independent of itself then either $P(A) = 0$ or $P(A) = 1$.

(b) Prove that if A is an event such that either $P(A) = 0$ or $P(A) = 1$ then A is independent of itself.

Ex. 1.4.9. This exercise explores the relationship between independence and conditional probability.

(a) Suppose A and B are independent events with $0 < P(B) < 1$. Prove that $P(A|B) = P(A)$ and that $P(A|B^c) = P(A)$.

(b) Suppose that A and B are independent events. Prove that A and B^c are also independent.

(c) Suppose that A and B are events with $P(B) > 0$. Prove that if $P(A|B) = P(A)$, then A and B are independent.

(d) Suppose that A and B are events with $0 < P(B) < 1$. Prove that if $P(A|B) = P(A)$, then $P(A|B^c) = P(A)$ as well.

Ex. 1.4.10. In this section we mentioned the following theorem: "If $E_1, E_2, \ldots, E_n$ is a collection of mutually independent events, then any subcollection of these events is mutually independent". Follow the steps below to prove the theorem.

(a) Suppose A, B, and C are mutually independent. In particular, this means that

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \text{ and}$$

$$P(A \cap B \cap C^c) = P(A) \cdot P(B) \cdot P(C^c).$$

Use these two facts to conclude that A and B are pairwise independent.

(b) Suppose $E_1, E_2, \ldots, E_n$ is a collection of mutually independent events. Prove that $E_1, E_2, \ldots, E_{n-1}$ is also mutually independent.

(c) Use (b) and induction to prove the full theorem.

1.5 USING R FOR COMPUTATION

As we have already seen, and will see throughout this book, the general approach to solve problems in probability and statistics is to put them in an abstract mathematical framework. Many of these problems eventually simplify to computing some specific numbers. Usually these computations are simple and can be done using a calculator. For some computations however, a more powerful tool is needed. In this book, we will use a software called R to illustrate such computations. R is freely available open source software that runs on a variety of computer platforms, including Windows, Mac OS X, and GNU/Linux.

R is many different things to different people, but for our purposes, it is best to think of it as a very powerful calculator. Once you install and start R,[1] you will be presented with a prompt that looks like the "greater than" sign (>). You can type expressions that you want to evaluate here and press the Enter key to obtain the answer. For example,

```
> 9 / 44
[1] 0.2045455
> 0.6 * 0.4 + 0.3 * 0.6
[1] 0.42
> log(0.6 * 0.4 + 0.3 * 0.6)
[1] -0.8675006
```

[1] Visit https://www.r-project.org/ to download R and learn more about it.

It may seem odd to see a `[1]` at the beginning of each answer, but that is there for a good reason. R is designed for statistical computations, which often require working with a collection of numbers, which following standard mathematical terminology are referred to as *vectors*. For example, we may want to do some computations on a vector consisting of the first 5 positive integers. Specifically, suppose we want to compute the squares of these integers, and then sum them up. Using R, we can do

```
> c(1, 2, 3, 4, 5)^2

[1]  1  4  9 16 25

> sum(c(1, 2, 3, 4, 5)^2)

[1] 55
```

Here the construct `c(...)` is used to create a vector containing the first five integers. Of course, doing this manually is difficult for larger vectors, so another useful construct is `m:n` which creates a vector containing all integers from `m` to `n`. Just as we do in mathematics, it is also convenient to use symbols (called "variables") to store intermediate values in long computations. For example, to do the same operations as above for the first 40 integers, we can do

```
> x <- 1:40
> x

 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22
[23] 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

> x^2

 [1]    1    4    9   16   25   36   49   64   81  100  121  144  169
[14]  196  225  256  289  324  361  400  441  484  529  576  625  676
[27]  729  784  841  900  961 1024 1089 1156 1225 1296 1369 1444 1521
[40] 1600

> sum(x^2)

[1] 22140
```

We can now guess the meaning of the number in square brackets at the beginning of each line in the output: when R prints a vector that spans multiple lines, it prefixes each line by the index of the first element printed in that line. The prefix appears for scalars too because R treats scalars as vectors of length one.

In the example above, we see two kinds of operations. The expression `x^2` is interpreted as an element-wise squaring operation, which means that the result will have the same length as the input. On the other hand, the expression `sum(x)` takes the elements of a vector `x` and computes their sum. The first kind of operation is called a *vectorized operation*, and most mathematical operations in R are of this kind.

To see how this can be useful, let us use R to compute factorials and binomial coefficients, which will turn up frequently in this book. Recall that the binomial coefficient

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

represents the number of ways of choosing k items out of n, where for any positive integer m, $m!$ is the product of the first m positive integers. Just as `sum(x)` computes the sum of the elements of `x`, `prod(x)` computes their product. So, we can compute $10!$ and $\binom{10}{4}$ as

```
> prod(1:10)
[1] 3628800
> prod(1:10) / (prod(1:4) * prod(1:6))
[1] 210
```

Unfortunately, factorials can quickly become quite big, and may be beyond R's ability to compute precisely even for moderately large numbers. For example, trying to compute $\binom{200}{4}$, we get

```
> prod(1:200)
[1] Inf
> prod(1:200) / (prod(1:4) * prod(1:196))
[1] NaN
```

The first computation yields `Inf` because at some point in the computation of the product, the result becomes larger than the largest number R can store (this is often called "overflowing"). The second computation essentially reduces to computing `Inf/Inf`, and the resulting `NaN` indicates that the answer is ambiguous. The trivial mathematical fact that

$$\log m! = \sum_{i=1}^{m} \log i$$

comes to our aid here because it lets us do our computations on much smaller numbers. Using this, we can compute

```
> logb <- sum(log(1:200)) - sum(log(1:4)) - sum(log(1:196))
> logb
[1] 17.98504
> exp(logb)
[1] 64684950
```

R actually has the ability to compute binomial coefficients built into it.

```
> choose(200, 4)
[1] 64684950
```

These named operations, such as `sum()`, `prod()` `log()`, `exp()`, and `choose()`, are known as *functions* in R. They are analogous to mathematical functions in the sense that they map some inputs to an output. Vectorized functions map vectors to vectors, whereas summary functions like `sum()` and `prod()` map vectors to scalars. It is common practice in R to make functions vectorized whenever possible. For example, the `choose()` function is also vectorized:

```
> choose(10, 0:10)
 [1]   1  10  45 120 210 252 210 120  45  10   1
> choose(10:20, 4)
 [1]  210  330  495  715 1001 1365 1820 2380 3060 3876 4845
> choose(2:15, 0:13)
```

```
[1]   1   3   6  10  15  21  28  36  45  55  66  78  91 105
```

A detailed exposition of R is beyond the scope of this book. In this book, we will only use relatively basic R functions, which we will introduce as and when needed. There are many excellent introductions available for the interested reader, and the website accompanying this book (TODO) also contains some supplementary material. In particular, R is very useful for producing statistical plots, and most figures in this book are produced using R. We do not describe how to create these figures in the book itself, but R code to reproduce them is available on the website.

EXERCISES

Ex. 1.5.1. In R suppose we type in the following

```
> x <- c(-15, -11, -4, 0, 7, 9, 16, 23)
```

Find out the output of the built-in functions given below:

```
sum(x)  length(x)  mean(x)  var(x)  sd(x)  max(x)  min(x)  median(x)
```

Ex. 1.5.2. Obtain a six-sided die, and throw it ten times, keeping a record of the face that comes up each time. Store these values in a vector variable `x`. Find the output of the built-in functions given in the previous exercise when applied to this vector.

Ex. 1.5.3. Use R to verify the calculations done in Example 1.2.4.

Ex. 1.5.4. We return to the Birthday Problem given in Exercise 1.2.12. Using R, calculate the Probability that at least two from a group of N people share the same birthday, for $N = 10, 12, 17, 26, 34, 40, 41, 45, 75, 105$.