Chapter 2

Probability

- Learning objectives:
  - understand how to determine the sample space in a random experiment
  - be familiar with events and set operations performed on events
  - understand the definition of and how to use discrete probability functions, including the equally likely probability function
  - be familiar with some of the basic rules of probability
  - understand various counting techniques involving permutation, combinations, and distinguishable permutations
  - be able to compute conditional probabilities, as well as use Bayes' Theorem
  - understand the definition of independent events and its implications

It is hard to imagine that anybody has not encountered probability in some form at some point in their life. That is, we all have some notion about what probability is. We feel that probability means the "chance that something happens". This, of course, is not a definition and is really only another way to say the probability that something happens.

Certainly, most of us have some comfort level with probability and even feel that we know certain things. But if we were asked to write out what it is that we know, our paper might not contain much writing when we are done. That is, we feel that we know things, but are not exactly sure what it is that we know. In this chapter, we will use definitions and theorems to formalize probability mathematically. Additionally, we will provide the student with examples and conversation in an effort to build an intuitive feel for probability. We take the approach that solving problems should be a thought process and not a search for the proper formula.

2.1 Random Experiments – Sample Space, Outcomes, and Events

Prior to performing some specific act or experiment, we often cannot say with certainty what will take place. Some examples would be rolling a die or pair of dice, tossing a single coin or multiple coins, or measuring the length of time that it will take to drive to work tomorrow. These are just a few of the endless possibilities of experiments with unknown results. The results will be known after we perform our experiment.

Definition 2.1.1

A **random experiment** is an experiment for which the result cannot be determined in advance. We often say there is an element of chance as to what the result will be. Examples of random experiments were given above.

Definition 2.1.2
Given some random experiment, a possible result of the experiment is called an **outcome**.

Definition 2.1.3
The collection of all possible outcomes of a random experiment is the **sample space**. We will denote the sample space of an experiment with the symbol $\mathcal{S}$.

Example 2.1.1 (Experiment 1)
A single die is rolled and the number of spots on the top side is noted. While we know the outcome will be one of $1, 2, 3, 4, 5, 6$, we do not know which one. So the experiment is a random experiment as defined above. The set of all possible outcomes for this experiment yields the sample space:

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}.$$

In Experiment 1, the numbers $1, 2, 3, 4, 5, 6$ are the possible outcomes of the experiment. We note that there are six outcomes in the sample space.

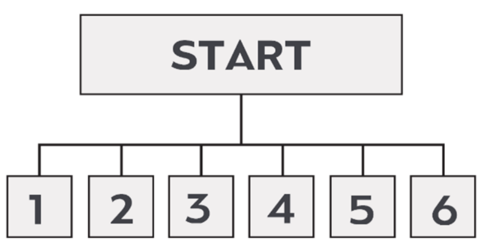
Example 2.1.2 (Experiment 2)
A coin is to be tossed three times. The sample space can be carefully constructed as

$$\mathcal{S} = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{HTT}, \text{THT}, \text{TTH}, \text{TTT}\}.$$

In this experiment, there are a total of $2^3 = 8$ outcomes in the sample space.

Example 2.1.3 (Experiment 3)
Two dice are to be rolled. Can you guess how many outcomes are in the sample space?
Your Guess: ____________ Actual: ____________

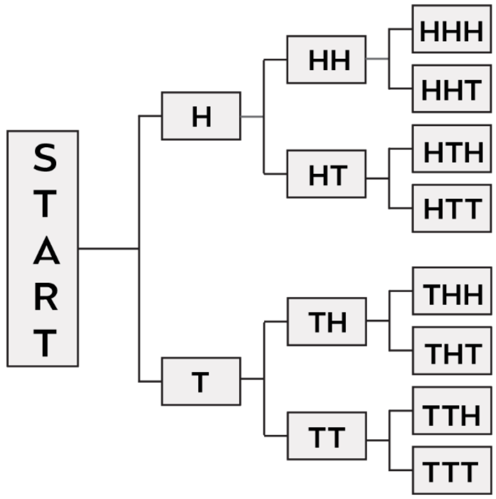
It is sometimes helpful to draw a tree diagram in order to determine or count all of the possible outcomes of an experiment. This is especially helpful when we have an experiment that takes place in stages, like tossing three coins. A tree diagram is a diagram that reveals the possible branches/outcomes at each stage of the experiment. It would be rather boring and not very enlightening to draw a tree diagram of Experiment 1, but it can be done:



At the start of the experiment, there was nothing. We roll the die, and we are done. The sample space of outcomes consists of just the six possible results: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

If we were to roll the die a second time, as in Experiment 3, there would be 6 branches from each of the 6 first-roll outcomes. There would be a total of 36 outcomes that can be written in the form of ordered pairs, such as $(1, 1)$ or $(3, 5)$.

A tree diagram for Experiment 2 would involve a two-branch break at each coin toss. At the start of the experiment, we had nothing. We now toss the coin. We have two branches corresponding to the two possible outcomes on the first toss: H or T. From each of these branches, we toss the coin again. This will create two new branches from each of the previous branches. This gives four possibilities after two tosses: HH, HT, TH, or TT. After tossing the coin for a third time, we have two new branches emanating from each of the previous four. Our tree diagram is now complete, and we can easily see the 8 outcomes in the sample space.



> **Example 2.1.4** (Experiment 4)
>
> Suppose that a coin is to be tossed until a heads is observed. Once a heads occurs, the experiment is terminated. Determine the sample space and draw the tree diagram.

In this last experiment, the sample size contains an infinite number of possible outcomes, but still can be listed. Any set that has a finite number of outcomes (like the sample spaces for Experiments 1-3) or can be listed (like the sample space for Experiment 4) is called a **discrete set**. Another name for a set that is discrete is a **countable set**.

> **Example 2.1.5** (Experiment 5)
>
> A coin is to be tossed. If the outcome is heads, the experiment ends. If the outcome is tails, we roll a standard six-sided die. Determine the sample space.

Example 2.1.6 (Experiment 6)

Suppose that a company is making coins that are supposed to weigh 3 grams. For a variety of reasons, the coin-making process includes some amount of variation, so that the weights of the coins produced will vary. The actual weights of the coins will not be exactly 3 grams, but close to 3 grams. Assume that all of the coins will be within 0.1 grams of the target value 3 grams. The sample space of possible weights cannot be listed because it is an interval of numbers. In this experiment, the data can take on any value between 2.9 and 3.1, and we write the sample space as an interval: $\mathcal{S} = [2.9, 3.1]$.

The set $[2.9, 3.1]$ that occurred as the sample space in Experiment 6 is not a discrete set. This type of set is called an **interval set**. Sets like this have so many elements that they cannot be put into a list. In the next section, we will discuss how many outcomes are in a set or in a sample space. Our initial focus will be on experiments with a discrete sample space.

Definition 2.1.4

Given some random experiment and the sample space $\mathcal{S}$, an **event** is defined to be any collection of outcomes from $\mathcal{S}$.

Note that the sample space itself is an event since it is a collection of outcomes. Additionally, the empty set (denoted by $\emptyset$) is an event, but it is an event that has no outcomes. Below we list some examples of events based on the experiments considered earlier.

Example 2.1.7 (Experiment 1)

A single die is rolled. Some examples of events include:

- E_1 = An odd number was rolled = $\{1, 3, 5\}$
- E_2 = An even number was rolled = $\{2, 4, 6\}$
- E_3 = A three was rolled = $\{3\}$ (note that outcomes are also considered events)

Example 2.1.8 (Experiment 2)

A coin is to be tossed three times. Some examples of events include:

- E_1 = An odd number of heads were tossed = $\{\text{HHH}, \text{HTT}, \text{THT}, \text{TTH}\}$
- E_2 = No tails were tossed = $\{\text{HHH}\}$
- E_3 = Less than two heads were tossed = $\{$____________________$\}$

Example 2.1.9 (Experiment 3)

Two dice are to be rolled. Some examples of events include:

- E_1 = The two dice match = $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
- E_2 = The sum of the two dice equals 7 = $\{(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)\}$
- E_3 = The sum of the two dice is more than 10 = $\{$____________________$\}$
- $E_4 = \emptyset$ (of course, this event cannot occur but is still considered an event)
- $E_5 = \{(1, 6), (1, 7)\}$ (this is <u>not</u> considered an event because it contains something that is not a possible outcome)

Example 2.1.10 (Experiment 4)

Suppose that a coin is to be tossed until a heads is observed. Some examples of events include:

- E_1 = The experiment ended on the fifth toss = $\{\text{TTTTH}\}$
- E_2 = The experiment took more than 3 tosses = $\{\text{TTTH}, \text{TTTTH}, \text{TTTTTH}, \ldots\}$

- E_3 = The experiment ended after an even number of tosses = {__________________________}

Example 2.1.11 (Experiment 6)

Suppose that a company is making coins as described above. Some examples of events include:
- E_1 = The coin weighs less than 3.05 grams = $[2.9, 3.05]$
- E_2 = The coin weighs more than 3.2 grams = _____________

2.2 Events – Cardinality, Unions, Intersections, and Complements

We have seen that events are sets whose elements are outcomes from a random experiment. In order to work with events, we must be familiar with a variety of set operations, which we review in this section.

2.2.1 Events and Cardinality

A set is a collection of distinct objects or elements. The sample space of a random experiment is a set of importance in this text and in probability, in general. Recall that the sample space is the collection (or set) of all possible outcomes of an experiment. Therefore, in this setting, we will refer to the distinct objects or elements in a set as outcomes. Recall that an event is a set whose elements are outcomes from a random experiment. In other words, an event can be viewed as a subset of the sample space. Recall that a single outcome of a random experiment is also considered an event (i.e., it is a set containing one element).

In some of the examples in the previous section, we determined the number of outcomes in the sample space. This is often a quantity of great interest, as it represents the total number of things that can occur in a random experiment. A related question is how to determine the number of outcomes in a given event. In some cases, the answer is obvious. But in many cases, we need to rely on our intuition and familiarize ourselves with counting techniques.

Clearly, an event has as few as 0 outcomes (recall the empty set is considered an event). Also, an event cannot have more outcomes than the sample space. We now loosely define a measure of how big a set (or event) is.

Definition 2.2.1

The **cardinality** of the sample space or an event is a measure of the number of outcomes in the set. We denote the cardinality of the sample space or an event using notation like $C(\mathcal{S})$ or $C(E)$.

Definition 2.2.2

Two events are said to have the **same cardinality** if there exists a one-to-one and onto function, known as a bijection, from one event to the other.

Definition 2.2.3

Given the set of integers $\mathcal{S} = \{1, 2, 3, \ldots, n\}$, we define the cardinality of the set to be n. That is, $C(\mathcal{S}) = n$.

Example 2.2.1

$E_1 = \{1, 2, 3\}$ and $E_2 = \{3, 4, 5\}$ have the same cardinality since $f(x) = x + 2$ is a bijection (one-to-one and onto function) from E_1 to E_2. We write $C(E_1) = 3 = C(E_2)$.

Example 2.2.2

If our events do not consist of numbers, then the bijection will not be mathematical in nature. For instance, let $E_1 = \{A, B, C\}$ and $E_2 = \{\text{Chair,Table,Lamp}\}$.

We define $f(A) = \text{Chair}$, $f(B) = \text{Table}$, and $f(C) = \text{Lamp}$. This is a bijection from E_1 to E_2, and so the two events have the same cardinality. Of course, this is not a numerical function, but that is not a requirement.

It seems obvious that the cardinality of both events should be 3. However, we have really only defined the cardinality of the set of integers $\{1, 2, \ldots, n\}$, so to see that the cardinality of these events is 3, we should define a bijection between one of the events and the set $\{1, 2, 3\}$.

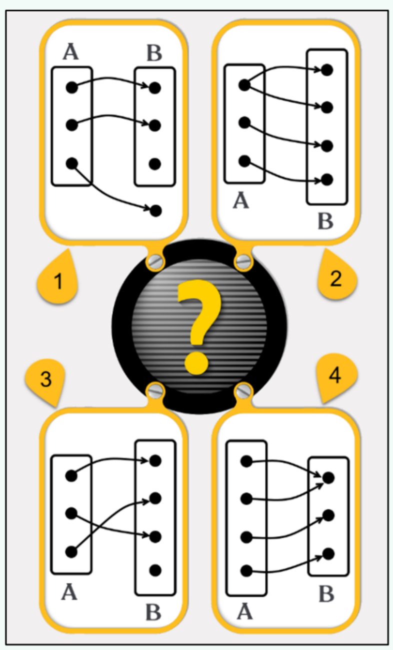
For instance, consider the function $g(A) = 1$, $g(B) = 2$, and $g(C) = 3$. This defines a bijection between E_1 and the set $\{1, 2, 3\}$. Since the latter has cardinality 3, we see that $C(E_1) = 3$. And since we have shown above that E_1 and E_2 have the same cardinality, we can also write $C(E_2) = 3$.

Example 2.2.3

Consider the events $E_1 = \{A, B, C, D\}$ and $E_2 = \{\text{Chair,Table,Lamp}\}$. Clearly, any function that maps all of E_1 onto E_2 cannot be one-to-one. So E_1 and E_2 do not have the same cardinality. In fact, $C(E_2) = 3$ from above, and it can be shown that $C(E_1) = 4$ using similar methods to those used in the previous example.

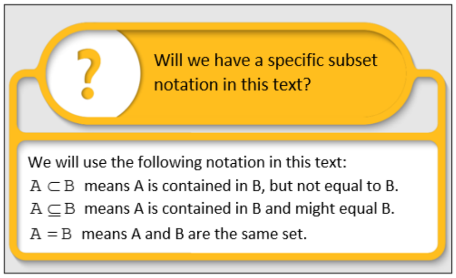
Example 2.2.4

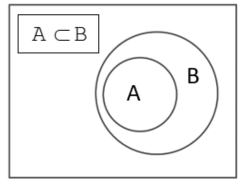
For each function shown in the image below, determine if f is a bijection from A to B.



Example 2.2.5

Consider the events $E_1 = \{1, 2, 3, \ldots\}$ and $E_2 = \{2, 4, 6, \ldots\}$. At first glance, the first event seems to have a larger cardinality since $E_2 \subset E_1$. However, in order for two events to have the same cardinality, we only need to find a bijection between the two sets. Notice that $f(x) = 2x$ defines a bijection from E_1 to E_2. Each outcome in E_1 maps into exactly one outcome in E_2, and each outcome in E_2 has exactly one outcome in E_1 mapped into it.





Definition 2.2.4

If an event has a finite cardinality or has the same cardinality as the natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$, the event is said to be **countable**, **countably infinite**, or **discrete**.

Definition 2.2.5

An infinite set that that is not countable is said to be **uncountable**.

Example 2.2.6

The set of all rational numbers in the interval $[0, 1]$ is countable.

The set of all real numbers, rational and irrational, in the interval $[0, 1]$ is uncountable.

The set of all irrational numbers in the interval $[0, 1]$ is uncountable.

2.2.2 Unions, Intersections, and Complements

Unions, intersections, and complements are concepts that should already be familiar to you. The definitions and examples are provided here anyway.

Definition 2.2.6

The **union** of two sets (events) is the collection of outcomes that are either in the first set or in the second set (or both). The notation for the union of two events is $A \cup B$. The union of two sets can easily be determined by taking the first set and adding the elements of the second set not already in the first set. In other words, the union combines the two sets (but any element that appears in both sets is only listed once).

Example 2.2.7

If $A = \{1, 2, 3\}$ and $B = \{2, 3, 4, 5\}$, then the union is $A \cup B = \{1, 2, 3, 4, 5\}$. The union is the collection of outcomes that are in at least one of the sets.

Definition 2.2.7

The **intersection** of two events is the collection of outcomes that the two events have in common. That is, the intersection is the collection of outcomes that are in the first event and also in the second event (in both). The notation for intersection of two events is $A \cap B$.

Example 2.2.8

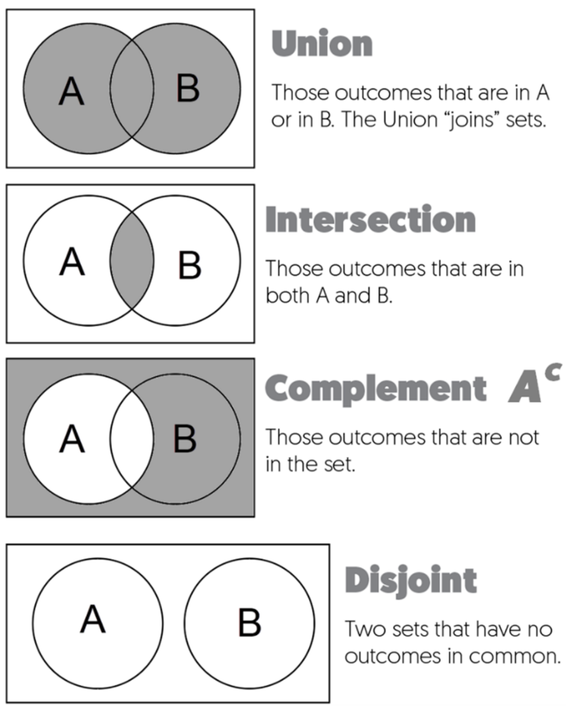
If $A = \{1, 2, 3\}$ and $B = \{2, 3, 4, 5\}$, then the intersection is $A \cap B = \{2, 3\}$. The intersection is the collection of outcomes that are in both of the sets.

Definition 2.2.8

The **complement** of an event consists of those outcomes in the sample space that are not in the event. The notation for the complement of an event E is E^C (other resources may use different notation). It should be noted that the complement can only be determined if the sample space is known.

Example 2.2.9

If $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ and $E = \{2, 3\}$, then $E^C = \{1, 4, 5, 6\}$. Note that if we did not know the sample space $\mathcal{S}$ (the collection of all possible outcomes), then we could not determine E^C.



Definition 2.2.9

Two events are said to be **disjoint** (or **mutually exclusive**) if they have no outcomes in common. That is, two events are disjoint if their intersection is empty. For instance, the sets $A = \{2, 3\}$ and $B = \{5, 6\}$ are disjoint because there are no outcomes that are in both sets. The sets $C = \{2, 3, 4\}$ and $D = \{4, 5, 6\}$ are not disjoint because the outcome 4 is in both sets.

2.3 Discrete Probability Functions

Definition 2.3.1

Let $\mathcal{S}$ be a discrete sample space from some experiment. A function P, defined on all events that consist of outcomes in $\mathcal{S}$, is said to be a **probability function** if it satisfies the following conditions:

(a) For any event E, $P(E) \geq 0$.

(b) $P(\mathcal{S}) = 1$

(c) If $E = \bigcup_{i=1}^{\infty} A_i$, where the events A_i are mutually exclusive, then $P(E) = \sum_{i=1}^{\infty} P(A_i)$.

The pair $\Omega = (\mathcal{S}, P)$ is said to be a **probability space**. In other words, a probability space is a sample space along with an associated probability function.

Remember: The input into P is an event. The output is a number in the interval [0,1]

Note that, at this time, our probability function is nothing more than a function on a sample space. The probability function does not represent the "chance" of an event. We will soon be looking at functions P that do represent our perception of chance.

It should be noted that in some texts, prior to defining a probability function, we would define the **event space** to be the collection of all possible subsets of $\mathcal{S}$ (often called the power set of $\mathcal{S}$ and denoted $\mathcal{F}$). We then define our probability function P on the event space $\mathcal{F}$. We would write the probability space as a tripe $\Omega = (\mathcal{S}, \mathcal{F}, P)$, i.e., the sample space $\mathcal{S}$, the set of all events that consist of outcomes in $\mathcal{S}$, and and a function on this set of events that satisfies the properties in the definition above.

Note 2.3.1

We will use the notation $P(E)$ when working with a probability function. Other textbooks may use notation like $Pr(E)$ or $\mathbb{P}(E)$.

Example 2.3.1

Suppose that we roll a 6-sided die. The sample space is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

Note that there are $2^6 = 64$ possible events that can be made from our sample space. That is $C(\mathcal{F}) = 64$. While any probability function must be defined on $\mathcal{F}$, we can use its definition on the individual outcomes of $\mathcal{S}$ in order to determine whether or not a function could serve as a probability function.

s	$P_1(s)$	$P_2(s)$	$P_3(s)$	$P_4(s)$	$P_5(s)$	$P_6(s)$	$P_7(s)$
1	0.2	0.1	0.2	0.1	$\frac{1}{6}$	0.1	0.1
2	0.2	0.1	0.1	0.1	$\frac{1}{6}$	0.1	0.1
3	0.2	0.1	-0.1	0.1	$\frac{1}{6}$	0.2	0.1
4	0.2	0.1	0.2	0.1	$\frac{1}{6}$	0.2	0.1
5	0.2	0.1	0.3	0.1	$\frac{1}{6}$	0.25	c
6	0.2	0.1	0.3	0.5	$\frac{1}{6}$	c	$0.85 - 2c$

(a) Determine which of P_1, P_2, P_3, P_4, P_5 are probability functions.

(b) If we decide to assign $P_5(\{1,2\}) = 0.4$, is P_5 be a probability function?
(c) What value of c (if any) makes P_6 a probability function?
(d) What value of c (if any) makes P_7 a probability function?

Theorem 2.3.1
For any event E, $P(E) \leq 1$.

Proof: Note that $\mathcal{S} = E \cup E^C$. Since E and E^C are disjoint, we have

$$1 = P(\mathcal{S}) = P(E) + P(E^C).$$

This gives $P(E) = 1 - P(E^C)$, which implies $P(E) \leq 1$ (since $P(E^C) \geq 0$). □

In many textbooks, the first statement in the definition of a probability function is written as "for any event E, $0 \leq P(E) \leq 1$". The latter part of the inequality is unnecessary in our definition due to this simple theorem.

Note 2.3.2
In what follows, proofs are only provided when easy to follow and instructive. Harder proofs or those that are lacking in intuition are omitted for the sake of time.

Theorem 2.3.2 The Complement Rule
For any event E, $P(E^C) = 1 - P(E)$.

Proof: Refer to the proof of the previous theorem. □

Example 2.3.2
Suppose we accept as truth that when drawing a single card from a deck of cards, the probability the selected card is an ace is $P(\text{Ace}) = \frac{4}{52}$. Then the probability of not selecting an ace is $P(\text{Not an Ace}) = \frac{48}{52}$.

2.4 Some Rules of Probability

We have already seen our first two rules (non-definition) of probability: for any event E, (i) $0 \leq P(E) \leq 1$ and (ii) $P(E^C) = 1 - P(E)$. In this section, we explore some additional rules of probability.

Recall that the definition of a probability function implies that $P(A \cup B) = P(A) + P(B)$ for disjoint sets A and B.

Example 2.4.1

Suppose we accept as truth that when drawing a single card from a deck of cards, the probability the selected card is an ace is $P(\text{Ace}) = \frac{4}{52}$. Suppose we also assume the same is true for a king, i.e., $P(\text{King}) = \frac{4}{52}$. Then the probability of selecting an ace or a king is $P(\text{Ace} \cup \text{King}) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52}$.

A common mistake in probability is to assume that $P(A \cup B) = P(A) + P(B)$ is always true. Note that this is not part of our definition of a probability function, unless there is a requirement that A and B are disjoint sets.

Example 2.4.2

Suppose we will draw a single card from a deck of cards. Let A be the set of red cards and let B be the set of diamonds. We will assign $P(A) = 0.5$ and $P(B) = 0.25$. Since B is a subset of A, it should be clear that $A \cup B = A$, so that $P(A \cup B) = P(A) = 0.5 \neq 0.75 = P(A) + P(B)$.

Learning to do this sort of investigative thinking will serve you well in any mathematics or statistics course.

Example 2.4.3

Let the sets A and B both be equal to $\mathcal{S}$ (the sample space). Note that $A \cup B = \mathcal{S} \cup \mathcal{S} = \mathcal{S}$. Again, we see that $P(A \cup B) = P(A) + P(B)$ is violated, as this leads to $P(\mathcal{S} \cup \mathcal{S}) = P(\mathcal{S}) + P(\mathcal{S}) = 1 + 1 = 2$. There are at least two problems with this statement, since we already defined $P(\mathcal{S}) = 1$ and showed that the probability of any event is at most 1.

The reason the formula $P(A \cup B) = P(A) + P(B)$ failed in the previous two examples is that the sets A and B were not disjoint. Clearly, the formula does not apply in such cases, and so we need to work on a formula for $P(A \cup B)$ when A and B are not disjoint.

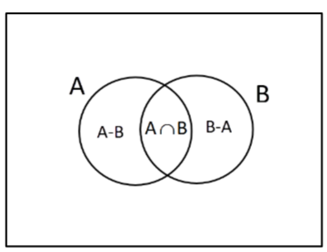
Definition 2.4.1

The **difference** of A and B, denoted by $A - B$, is the set of outcomes that are in A but not in B.

Similarly, we define the **difference** of B and A, denoted by $B - A$, to be the set of outcomes that are in B but not in A.

We can now write the union of two sets as a union of three disjoint sets as follows:

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



Using the addition rule for disjoint events, we get the following:

$$\begin{aligned} P(A \cup B) &= P(A - B) + P(A \cap B) + P(B - A) \\ &= P(A - B) + P(A \cap B) + P(B - A) + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

We now have our general rule for the probability of the union of two events.

Theorem 2.4.1
For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: Done above in the comments preceding the theorem. $\square$

Example 2.4.4
If $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.1$, determine $P(A \cup B)$.

Example 2.4.5
If $P(A) = 0.25$, $P(B) = 0.37$, and $P(A \cap B) = 0.12$, determine $P(A \cup B)$.

At this point, some of you may be considering a rule for intersections based on things that you remember from past courses or your intuition. A common guess for a formula that deals with the probability of an intersection is $P(A \cap B) = P(A) \cdot P(B)$. Is this statement true? Let's do some investigative thinking!

Theorem 2.4.2
If the events A and B are disjoint, then $P(A \cap B) = 0$.

Proof: Note that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ based on the previous theorem. But since A and B are disjoint, the third rule of a probability function gives that $P(A \cup B) = P(A) + P(B)$. In other words, it must be the case that

$$P(A) + P(B) - P(A \cap B) = P(A) + P(B),$$

which can only happen if $P(A \cap B) = 0$. $\square$

Theorem 2.4.3
If $A \subseteq B$, then $P(A \cap B) = P(A)$.

Proof: The general rule for unions gives that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Since $A \subseteq B$, notice that $A \cup B = B$. Substituting this into equation above gives $P(B) = P(A) + P(B) - P(A \cap B)$, so that $P(A \cap B) = P(A)$.

Alternatively, since $A \subseteq B$, we have that $A \cap B = A$ so that $P(A \cap B) = P(A)$. $\square$

The previous theorem implies that when it comes to the intersection of two events, $0 \leq P(A \cap B) \leq \min \{P(A), P(B)\}$.

Theorem 2.4.4
For any two events A and B, $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

Proof: Note that this is just an algebraic manipulation of the general union theorem from above. $\square$

It seems there are several ways to write the probability of an intersection of events, and it is seems unlikely that $P(A \cap B) = P(A) \cdot P(B)$ will be true, in general. We will need to revisit the probability of the intersection of two events later.

2.5 Probability Viewed as Relative Frequency

While we have defined a probability function mathematically as a function that follows a certain set of rules, we would like to put some practical meaning to probability. This might make studying probability more interesting. It would allow us to use this field of study to get a better understanding of real-world situations.

Suppose that we have a finite sample space consisting of n outcomes $O_1, O_2, \ldots, O_n$. We could make the following assignment of probability for each outcome in $\mathcal{S}$: $P(O_i) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$.

This is clearly a probability function on our sample space, and this function assigns an equal probability to each outcome in $\mathcal{S}$.

Example 2.5.1
Confirm the function defined above conforms to the requirements in our definition of a probability function.

Example 2.5.2
Suppose that we roll a single die. Since $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, the probability function described above assigns $P(O_i) = \frac{1}{6}$.

If we now think of probability as the "chance" that a specific outcome occurs, we might say that each outcome is "equally likely" to occur since each O_i has the same chance of occurring. That is, each outcome is equally as likely as any other outcome to occur. Based on our lifetime of experience using the words "chance", "probability", and "likelihood", we should start feeling more comfortable about this recently defined function called a probability function.

Note that a sample space does not need to have a finite number of outcomes. But it does need to be finite if we want to use equally likely outcomes in our probability function definition. A finite sample space with the assumption that outcomes are equally likely occurs often enough that the associated probability function deserves its own definition.

Definition 2.5.1

For a finite sample space consisting of n outcomes $O_1, O_2, \ldots, O_n$, the **equally likely probability function** is defined by $P(O_i) = \frac{1}{n}$ for $i = 1, 2, \ldots, n$.

Note 2.5.1

Just because a sample space has a finite number of outcomes and it is legal to assign $P(O_i) = \frac{1}{n}$, it may be an unwise assignment of probability if the assumption of equal likely outcomes seems invalid.

For example, consider purchasing a lottery ticket for an upcoming lottery. Since you will either win the lottery or not, the sample space of outcomes for you could be written as $\mathcal{S} = \{\text{I win}, \text{I do not win}\}$. Since the sample space is finite with $C(\mathcal{S}) = 2$, the equally likely probability function would lead to $P(\text{I win}) = \frac{1}{2}$ and $P(\text{I do not win}) = \frac{1}{2}$.

While legal under the rules of probability functions, it's clear that the probabilities assigned here disagree tremendously with the interpretation of probability as the "chance" that something occurs.

Nevertheless, at this point, we have only formally named one probability function (equally likely outcomes), and so we are not currently in a position to choose between multiple functions. Choosing between probability functions will be clarified in later chapters as new situations arise. At this point it's enough to keep in mind that we may assign different probability functions other than the "equally likely" probability function.

Example 2.5.3

What sample space considered earlier is a setting where assigning equal probability to all outcomes would be a poor choice? (Even though it would be legal, it goes against our notion of "chance".)

Besides computing probabilities of individual outcomes in the sample space, we can use our knowledge of events to calculate the probability of a collection of outcomes. A strategy here is to write the event in question in terms of outcomes in the sample space and, only then, apply the probability function.

Note 2.5.2

When asked to compute the probability of an event using the equally likely probability function, re-write the event in terms of disjoint events joined together by "or", and then compute the probability by summing the probabilities of each of the "or" events.

Suppose that we have a standard deck of 52 cards. If we are going to select a card at random from the deck, it's reasonable to assume that each card has an equal chance of being selected. Therefore, a reasonable assignment of probability to each card would be $\frac{1}{52}$, i.e., each card is equally likely to be selected, and so we apply the equally likely probability function to the 52-card sample space.

Example 2.5.4

Let A be the event that a heart is selected. Since each card is disjoint from every other card, we can determine the probability of the event A as follows. First, express the event A in terms of elements of the sample space:

$$A = \text{Heart is selected} = \{2\heartsuit, 3\heartsuit, 4\heartsuit, \ldots, 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, A\heartsuit\}$$

Then apply the equally likely probability function as follows:

$$\begin{aligned}
P(A) &= P(\text{Heart is selected}) \\
&= P(\{2\heartsuit, 3\heartsuit, 4\heartsuit, \ldots, 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, A\heartsuit\}) \\
&= P(2\heartsuit) + P(3\heartsuit) + P(4\heartsuit) + \cdots + P(10\heartsuit) + P(J\heartsuit) + P(Q\heartsuit) + P(K\heartsuit) + P(A\heartsuit) \\
&= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \cdots + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} \\
&= \frac{13}{52} \\
&= \frac{1}{4}
\end{aligned}$$

Note that we utilize the knowledge that each event in the first equation is mutually exclusive, that is, we can't draw one card that is both $2\heartsuit$ and $3\heartsuit$. Hence, these mutually exclusive "or" statements can be added together to compute the probability.

Notice that what we ultimately ended up computing was $P(A) = \frac{C(A)}{C(\mathcal{S})}$, i.e., the probability of event A is the number of outcomes in A divided by the total number of outcomes in the samples space. This approach can be used in subsequent examples and is explored further in the next section on counting techniques.

Example 2.5.5

A single card is selected from a deck of cards. Let A be the event that a king is selected. Determine $P(A)$.

Example 2.5.6

A single card is selected from a deck of cards. Let A be the event that a red card is selected. Determine $P(A)$.

Example 2.5.7

A single card is selected from a deck of cards. Let A be the event that a red five is selected. Determine $P(A)$.

We would like to consider some more complicated experiments. We now consider experiments that contain two or more actions.

Example 2.5.8

Below are some experiments that consist of two or more actions:

- Draw two cards from a deck.
- Draw three cards from a deck.
- Toss a coin five times.
- Make a license plate that consists of 3 letters followed by 4 digits.
- Choose 5 balls from a bin of 54 balls numbered 1 through 54 without replacing a ball after it has been selected.
- Choose 4 balls from a bin of 10 balls numbered 0 through 9, where we will replace each ball after it has been selected.

In the examples above, it seems reasonable to assign equal probability to each outcome in the sample space. In other words, the equally likely probability function seems like a good probability model to use here. This means that for each outcome $O_i \in \mathcal{S}$, we assign $P(O_i) = \frac{1}{n}$. Of course, to determine the numerical value of this probability assignment in a given situation, we need to be able to count the number of outcomes in the sample space $\mathcal{S}$. And as we saw in the examples preceding this discussion, if an event E consists of multiple (disjoint) outcomes from the sample space, we can determine the probability of the event using $P(E) = \frac{C(E)}{n}$. This also demonstrates the need to be able to count the number of outcomes in an event E. In some cases, determining $C(E)$ and $C(\mathcal{S})$ are straightforward (e.g., drawing a king from a deck of cards). In others, the counting can be a bit more challenging (e.g., randomly selecting a license plate that has no repeated letters or digits or being dealt a straight in a game of poker).

Note 2.5.3

Often times, the equally likely probability function seems logical to use. That is, the assignment $P(O_i) = \frac{1}{n}$ feels like the right way to describe the "chance" that O_i will occur. It is important to note that we can never prove that this function gives the true "chance" of each of the outcomes. This is just our model, and with this model we make decisions as if it were 100% accurate.