1 Probability Essentials

In general parlance, the term *probability* is used as a measure of a person's belief in the occurrence of an event. This is the *subjective* notion of probability. Take a sentence such as, "There is a 60% probability of rain tomorrow". Where did this come from? A professional meteorologist has used a sophisticated mathematical model to distill large amounts of atmospheric data down to a simple number indicating her belief that it is more likely to rain than to not rain tomorrow. We can intepret this number however we wish. We should be careful, however, to examine both our biases and those of the meterologist. Who does this meterologist work for? How reliable have their predictions been in the past? (We should be careful here since we might be more likely to recall rain when none was predicted than the other way around.) Are meterologists more likely to err on the side of predicting rain, since that way fewer people will be upset if they are wrong? We can use this number to make important life decisions such as whether or not to carry an umbrella. The downside of this is that it is in essence a subjective opinion, even if that opinion comes from an expert. In addition, for subjective probabilites to make sense, they have to be internally consistent. In this example, if we believe there is a 60% chance of rain tomorrow, we must also believe that there is a 40% chance of it not raining tomorrow.

Let's look at a different example. What is the probability of rolling a 1 on a standard six-sided die? We can argue based on symmetry that since dice are cubical, there should be no reason for one face to be preferred over another. Thus this probability is 1/6. This is the *classical* notion of probability. In this interpretation, we use symmetry arguments to divide our experiment (a single die roll, in this case) into elementary events which are equally probable (the six distint die rolls). We can compute the proability of any event using these elementary, equiprobably events. The limitation of this approach is that is requires a symmetry argument to be effective; thus it works for coin flips and die rolls, but not for more complicated scenarios.

We can look at the die roll experiment in another way. Imaging rolling a standard six-sided die repeatedly. The empirical probability of rolling a 1 is the ratio of the number of times a 1 is rolled to the total number of rolls. In general, we have:

$$\text{empirical probability of a certain event} = \frac{\text{number of times the event occurs}}{\text{total number of trials}}$$

Intuitively, as we perform more and more dice rolls, the empirical probability of rolling a 1 should approach some mythical quantity which we call the *true probability* of rolling a 1. This approach is the *empirical*, or *frequentist*, approach to probability. For a standard, six-sided die, it stands to reason that the empirical probabilty should approach the classical probabiliy of 1/6 as the number of rolls approaches infinity. This result is called the *law of large numbers* and will be discussed later in the course. The empirical approach has a critical advantage over the classical approach in that we do not require symmetry to compute our probabilities. As an example of this, think of how you would determine the probability of rolling a 1 if someone handed you a loaded die. The empirical approach does, however, have its limits. Returing to the weather example, there is no way to think of the chance of rain tomorrow as the limit of a sequence of independent experiements. Unless we are in the movie Groundhog Day, tomorrow can only happen once!

For our purposes, we require a more rigorous, mathematical construction of probability. This is known as *axiomatic probability* and will unify some of the aspects of the other approaches to probability. For this, we turn to the language of *set theory*.

1.1 Sample Spaces

A *set* is a collection of distinct objects. A *sample space*, denoted $\mathcal{S}$ is the set of all outcomes of a particular experiment. Here are some examples of sample spaces:

1. Single coin flip: $\mathcal{S} = \{H, T\}$
2. Roll of one standard, six-sided die: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$
3. Roll of two standard, six-sided dice: Here we represent the sample spaces as ordered pairs.

		second roll					
		1	2	3	4	5	6
	1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
first roll	3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

4. Number of free throw attempts it takes for me to make a single basket: $\mathcal{S} = \{1, 2, 3, ...\}$. This set is often denoted $\mathbb{N}$ for the natural numbers (positive integers)
5. Number of minutes late my RIPTA bus arrives: $\mathcal{S} = [0, \infty)$.

Note that the first three sample spaces contain only a finite number of elements (2, 6, and 36 elements, respectively). These are called *finite sample spaces*. The fourth and fifth sample spaces both contain an infinite number of elements, but there is a fundamental difference between the two. The set $\mathbb{N}$ can be written out in its entirety in an infinitely long list; another way to think about this is that we can start at 1 and count up to any number in the set (as long as we have enough time!). A set with this property is called *countable*. For the set $[0, \infty)$, it makes intuitive sense that we cannot do this, i.e. we cannot list all the elements and, say, "count up to π". A proof of this fact is left for another course. Such an infinite set is called *uncountable*[1]. A sample space which is either finite or countable is called *discrete*.

[1] An uncountable set is a "larger infinity" than a countable set, which leads to the concept of "sizes of infinity". John Green alludes to this in his novel *The Fault in Our Stars*, but unforunately gets the math wrong. If you find this interesting, I recommend the Vi Hart video `https://www.youtube.com/watch?v=23I5GS4JiDg`

1.2 Events and Subsets

An *event* is a subset of a sample space. Events are usually designated by capital letters, and we write the relationship "A is a subset of $\mathcal{S}$" by $A \subset \mathcal{S}$. For two events A and B, $A \subset B$ if every element in A is also contained in B. The *empty set*, denoted $\emptyset$, is the set containing no elements, and it is a subset of every set.

Let us consider the sample space $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, representing the roll of a single die. The following are examples of events:

1. $A = \{2, 4, 6\}$, the event that an even number is rolled
2. $B = \{1, 2, 3\}$, the event that the roll is less than or equal to 3
3. $C = \{1\}$, the event that a 1 is rolled

The event C consists of a single element in the sample space. Such an event is called a *simple event* and cannot be decomposed. The events A and B are each composed of three simple events.

Next, consider the sample space repesenting rolls of two dice. Let E be the event that the sum of the two dice is 7. We can represent this event graphically; in the figure below, the event E consists of the squares which are highlighted in yellow.

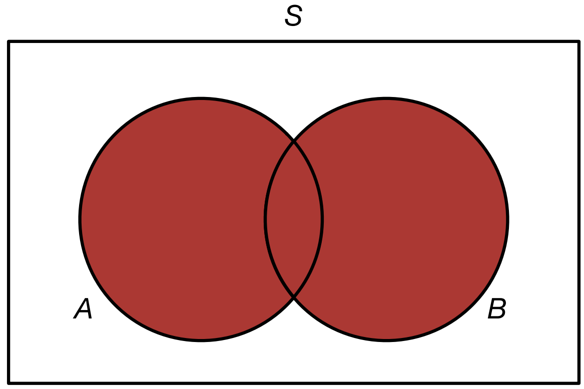
		second roll					
		1	2	3	4	5	6
	1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
first roll	3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

1.3 Basic Set Operations

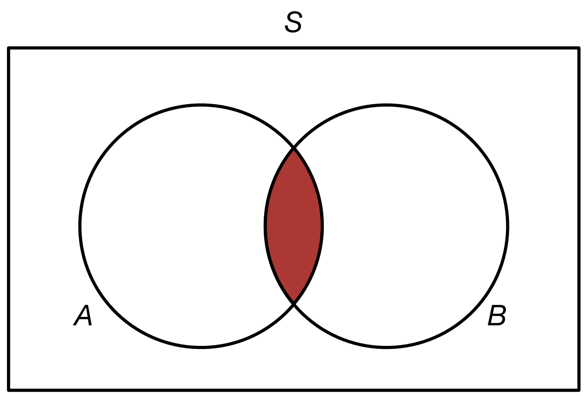
> "You have multiple core competencies with surprisingly minimal Venn. You can pivot from working on astrophysics problems, to teaching the young Arkers, to podcasting to folks on the ground, without skipping a beat!" - Neal Stephenson, *Seveneves*

Let $\mathcal{S}$ be our sample space, the set of all elements under consideration. Consider two events A and B which are subsets of $\mathcal{S}$. We have the following three basic set operations, which are handily illustrated using Venn diagrams.

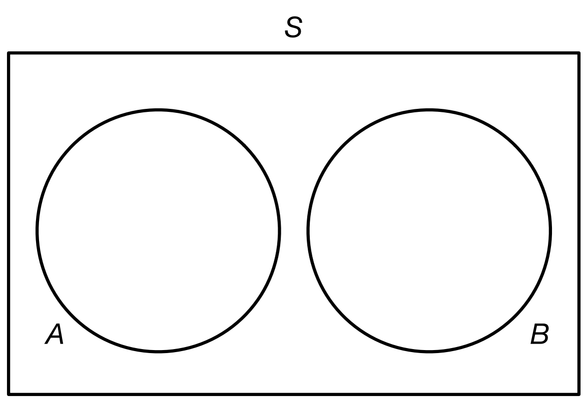
1. The *union* of A and B, denoted $A \cup B$, is the set of all elements which are in A or B (or both). That is, the union is all elements that are in at least one of the two sets.



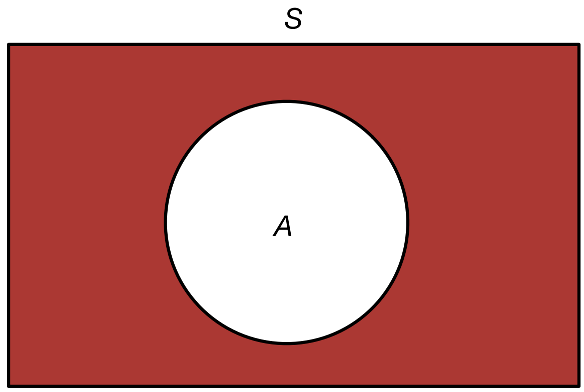
2. The *intersection* of A and B, denoted $A \cap B$, is the set of all elements which are in both A and B.



Two sets A and B are *disjoint* or *mutually exclusive* if they have no elements in common, i.e. if $A \cap B = \emptyset$.



3. The *complement* of A, denoted A^c, is the set of all points in $\mathcal{S}$ which are not in A. Note that A and A^c are disjoint, and $A \cup A^c = \mathcal{S}$.



There are many relationships between these operations which fall under the rubric of set algebra. Most of them we will not need, but we mention a few useful ones here:

1. Distributive laws

$$
\begin{aligned}
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)
\end{aligned}
$$

2. DeMorgan's laws

$(A \cap B)^c = A^c \cup B^c$ (the complement of an intersection is the union of the complements)

$(A \cup B)^c = A^c \cap B^c$ (the complement of a union is the intersection of the complements)

1.4 Axiomatic Definition of Probability

Equipped with our our knowlege of set theory, we can define probability axiomatically as follows. Given any event A in our sample space $\mathcal{S}$, we assign a probability $\mathbb{P}(A)$ to that event such that the following rules hold[2]:

1. $0 \leq \mathbb{P}(A) \leq 1$

   The probability of an event is a real number between 0 and 1, where a probability of 0 means that the event will never occur, and a probability of 1 means that the event will always occur.

2. $\mathbb{P}(\emptyset) = 0$

   The probability that nothing happens is 0, i.e. something must happen.

[2]You can construct a coherent notion of probabiltiy with fewer axioms and derive the remaining rules from these; I like this version of probability rules, since it codifies what we want to be true given our intuitive notion of probability.

3. $\mathbb{P}(\mathcal{S}) = 1$

   The probability of the whole sample space is 1, which is another way of saying that something must happen.

4. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

   If we make a set bigger, its probability can only increase (or stay the same); it cannot decrease.

5. If $A_1, A_2, \ldots, A_n$ are pairwise disjoint events, i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$, then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

   This holds for an infinite sequence as well, i.e. if $A_1, A_2, A_3, \ldots$ are a sequence of pairwise disjoint events, then

$$\mathbb{P}(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

From these we can derive a very important rule:

$$\mathbb{P}(A) + \mathbb{P}(A^c) = 1, \quad \text{i.e. } \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

Sometimes it is easier to calculate the probability of an event *not* happening than the probability of the event itself!

These rules tell us the properties that we want probability to have. However, given a sample space, they do not actually tell us how to assign probabilities to each event in the sample space. Doing that in a way that is consistent with the above rules can be a bit tricky [3], but luckily for a discrete sample space we can do this with no problem. Since a discrete sample space is composed of a finite (or countable) number of simple events, all we have to do is assign probabilities to each simple event in such a way that they all add up to 1.

Example. Consider once again tossing a single die. The sample space for this is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. This sample space contains 6 simple events: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$, and $\{6\}$. We can assign any probabilties we want to these simple events, as long as they add up to 1. For example, assuming we have a fair die, we can let $\mathbb{P}(\{i\}) = 1/6$ for $i = 1, \ldots, 6$. If we like, we can check that all the above rules hold. If we have a loaded die, which rolls a 6, say, half the time, we could assign probabilites: $\mathbb{P}(\{6\}) = 1/2, \mathbb{P}(\{i\}) = 1/10$ for $i = 1, \ldots, 5$.

[3] In fact, for an uncountable sample space such as $\mathcal{S} = [0, 1]$, you can show that you cannot construtct a notion of probability which is consistent with all the rules; this is the starting point for the development of measure theory, and is beyond the scope of this course.

Example. Consider this time a countable sample space $\mathcal{S} = \mathbb{N} = \{1, 2, 3, \dots\}$. One possibility is to assign probabilities $\mathbb{P}(\{i\}) = 1/2^i$ for $i = 1, 2, 3, \dots$, i.e. $\mathbb{P}(\{1\}) = 1/2$, $\mathbb{P}(\{2\}) = 1/4$, $\mathbb{P}(\{3\}) = 1/8$, etc. Perhaps you recall from calculus that this is a geometric series with first element 1/2 and common ratio 1/2, and so we know its sum is:

$$\sum_{i=1}^{\infty} \mathbb{P}(\{i\}) = \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1}{2}\frac{1}{1 - 1/2} = 1$$

Since the sum of the probabilities of all the simple events is 1, we are all set! If you have not seen this before, we will cover this in more detail when we discuss the geometric distribution. In the meantime, here is a nice picture to convince you that the sum is indeeed 1.