

# 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 interpret this number however we wish. We should be careful, however, to examine both our biases and those of the meteorologist. Who does this meteorologist 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 meteorologists 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 probabilities 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 distinct die rolls). We can compute the probability of any event using these elementary, equiprobable events. The limitation of this approach is that it 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. Imagine 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 probability should approach the classical probability 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. Returning to the weather example, there is no way to think of the chance of rain

tomorrow as the limit of a sequence of independent experiments. 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
first roll	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)
	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:  $\mathcal{S} = \{1, 2, 3, \dots\}$ . 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  $\pi$ ”. A proof of this fact is left for another course. Such an infinite set is called *uncountable*<sup>1</sup>. A sample space which is either finite or countable is called *discrete*.

---

<sup>1</sup>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 unfortunately 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 representing 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.

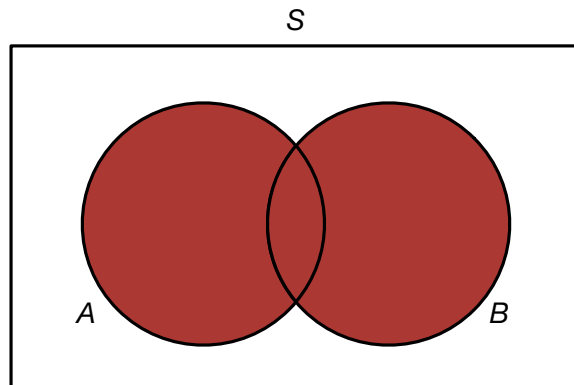
		second roll					
		1	2	3	4	5	6
first roll	1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
	2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
	3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
	4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
	5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
	6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6,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

## 1.3 Basic Set Operations

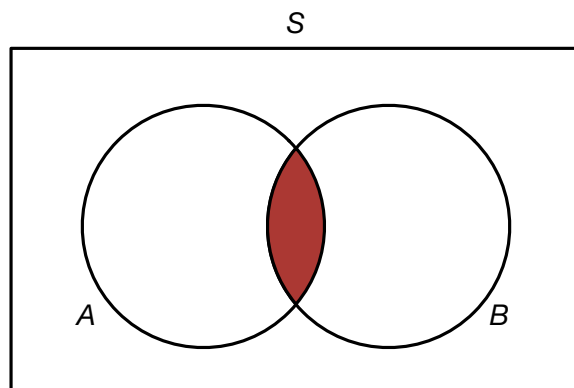
“You have multiple core competencies with surprisingly minimal Venn. You can pivot from working on astrophysics problems, to teaching the young Arkers, to podcasting to folks on the ground, without skipping a beat!” - Neal Stephenson, *Seveneves*

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.

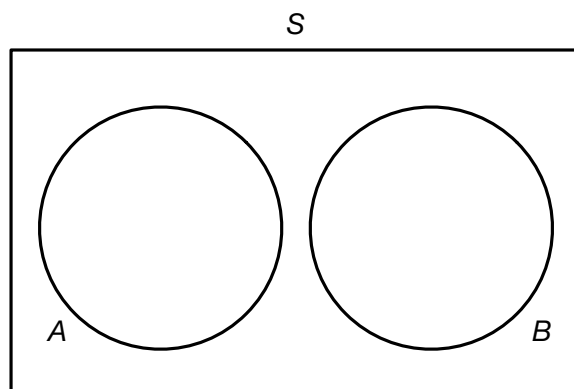
1. The *union* of  $A$  and  $B$ , denoted  $A \cup B$ , is the set of all elements which are in  $A$  or  $B$  (or both). That is, the union is all elements that are in at least one of the two sets.



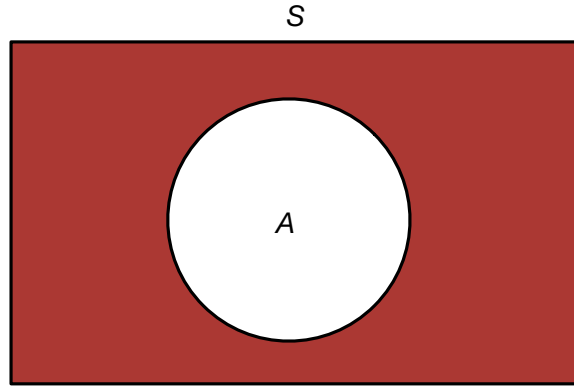
2. The *intersection* of  $A$  and  $B$ , denoted  $A \cap B$ , is the set of all elements which are in both  $A$  and  $B$ .



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  $S$  which are not in  $A$ . Note that  $A$  and  $A^c$  are disjoint, and  $A \cup A^c = S$ .



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

$$\begin{aligned} (A \cap B)^c &= A^c \cup B^c \quad (\text{the complement of an intersection is the union of the complements}) \\ (A \cup B)^c &= A^c \cap B^c \quad (\text{the complement of a union is the intersection of the complements}) \end{aligned}$$

## 1.4 Axiomatic Definition of Probability

Equipped with our knowledge 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<sup>2</sup>:

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.

---

<sup>2</sup>You can construct a coherent notion of probability 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  $A_1, A_2, \dots, A_n$  are pairwise disjoint events, i.e.  $A_i \cap A_j = \emptyset$  if  $i \neq j$ , then

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

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

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

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<sup>3</sup>, 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 probabilities 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  $i = 1, \dots, 6$ . 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 probabilities:  $\mathbb{P}(\{6\}) = 1/2, \mathbb{P}(\{i\}) = 1/10$  for  $i = 1, \dots, 5$ .

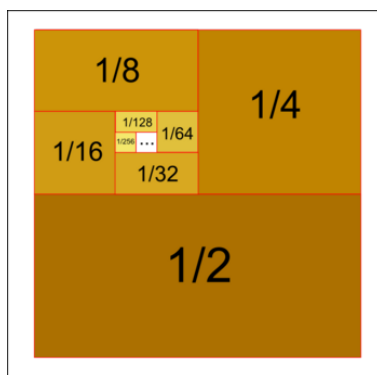
---

<sup>3</sup>In fact, for an uncountable sample space such as  $\mathcal{S} = [0, 1]$ , you can show that you cannot construct 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 indeed 1.



## 1.5 Discrete Uniform Distribution

The first probability distribution we will consider is the uniform distribution on a *finite* sample space. In the discrete uniform distribution, every simple event is equally likely to occur.

Suppose we have a finite sample space *calS* with  $n$  simple events. The discrete uniform distribution assigns each simple event a probability of  $1/n$ . Why is this the case? If each simple event has probability  $p$  and there are  $n$  simple events  $A_1, \dots, A_n$  in our sample space, then using the fact that the sample space  $S$  must have probability 1 and that the simple events are pairwise disjoint:

$$\begin{aligned} 1 &= \mathbb{P}(S) \\ &= \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n) \\ &= np \end{aligned}$$

where in the last line we use the fact that the simple events each have probability  $p$ , and there are  $n$  of them. Solving for  $p$ , we get  $p = 1/n$ .

How do we find the probability of an event that is not a simple event? Again, consider a discrete sample space  $\mathcal{S}$  with  $n$  simple events. Suppose we have an event  $A$  which is a subset

of  $\mathcal{S}$ . Since we have a finite sample space,  $A$  is composed of a finite number  $m$  of simple events, where  $m$  is an integer between 0 and  $n$ . The probability of  $A$  is:

$$\mathbb{P}(A) = \frac{\text{number of simple events in } A}{\text{number of simple events in } \mathcal{S}} = \frac{m}{n}$$

**Example.** Consider the finite sample space  $\mathcal{S}$  representing the roll of two standard dice. This sample space has 36 simple events, so each simple event has a probability of  $1/36$ . Note that for the simple events in this sample space, the order of the die rolls matters.  $(1, 6)$  and  $(6, 1)$  are two different events; even though both events contain the same die rolls, for the former, the 1 is rolled first, while for the latter, the 6 is rolled first.

1. What is the probability that the sum of the two dice is 7? Looking at the graphical depiction of this event above, we see that there are 6 simple events that give us a sum of 7. Thus the probability of a sum of 7 is  $6/36 = 1/6$ .
2. What is the probability that the sum of the two dice is less than 11. In this case, it is easier to compute the probability that the sum is 11 or greater and then subtract that probability from 1. (Why can we do this?) Let  $A$  be the event that the sum is 11 or greater.  $A$  is composed of 3 simple events:  $(6, 6)$ ,  $(6, 5)$ , and  $(5, 6)$ . Thus we have  $\mathbb{P}(A) = 3/36$ . The probability we want is:

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - 3/36 = 33/36$$

In the previous example, it is relatively straightforward to draw the sample space, so we can essentially compute any probability we want simply by looking at the picture and counting which boxes comprise our event of interest. For more complicated problems, this is not as easy. Consider the following example:

**Example.** A communication system consists of  $n$  antennas arranged in a line. Exactly  $m$  out of the  $n$  antennas are defective. The system is functional if no two consecutive antennas are defective. Assuming that each linear arrangement of the antennas is equally likely, what is the probability that the system will be functional?

For small values of  $n$  and  $m$ , say  $n = 4$  and  $m = 2$ , we can write out all of the possible configurations. Representing a functional antenna by 1 and a defective antenna by 0, there are exactly six linear arrangements:

1. 0 1 1 0
2. 0 1 0 1
3. 0 0 1 1
4. 1 0 0 1
5. 1 0 1 0
6. 1 1 0 0

Take a moment to convince yourself that these are the only possible configurations. Configurations 1, 2, and 5 are functional, so there are 3 functional configurations out of 6 total



configurations. So in this case, the probability that the system is functional is  $3/6$ .

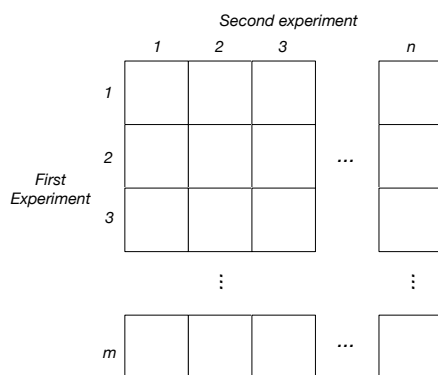
For general  $n$  and  $m$ , it is not immediately obvious how to perform the requisite counting of configurations. Taking a cue from the Count on Sesame Street, we need to learn more about counting. The mathematical theory of counting is known as *combinatorics*.

### 1.5.1 The Basic Principle of Counting

*The basic principle of counting (mn-rule)*

Suppose we are performing two experiments. If the first experiment has  $m$  possible outcomes and the second experiment has  $n$  possible outcomes, there are  $mn$  possible outcomes for the two experiments. We can generalize this to more than two experiments; in that case, we take the product of the number of outcomes from each experiment.

To see this, draw a grid of boxes with the  $m$  outcomes from the first experiment on the left and the  $n$  outcomes of the second experiment across the top. There are  $mn$  total boxes, which are all the possible outcomes of both experiments.



We have already seen the basic principle of counting in play when we looked at the sample space for two dice. Each of the two die rolls is a separate experiment. Since there are 6 outcomes for each die roll, the total number of outcomes is  $6 \cdot 6 = 36$ . We can extend this to three or more dice. For three dice, for example, the total number of outcomes is  $6 \cdot 6 \cdot 6 = 6^3 = 216$ . If you like, you can visualize this as a cube. For  $n$  dice, there are  $6^n$  possible outcomes. I have trouble visualizing this for  $n > 3$ , but perhaps you can.

We can also think of the basic principle of counting in terms of choosing items from groups. If there are  $m$  items in Group 1 and  $n$  items in Group 2, there are  $mn$  pairs of items consisting of one item from each group. Again, this can be extended to any number of groups.

**Example.** When I was growing up in Virginia, standard (non-vanity) license plates were composed of three letters (A-Z) followed by three digits (0-9). Assuming that all such possibilities can exist:

1. How many possible license plates are there?

We are choosing items from 6 groups. The first three groups contain 26 items, and the last three groups contain 10. Thus the number of possibilities is:  $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17,576,000$ . (That works out to a little more than two cars per person.)

2. If the Virginia DMV decided that letters and numbers could not be repeated, how many possible license plates would there be?

The difference here is that the group sizes shrink as we choose items. For the letters, the first group has 26 items. The second group only has 25 items, since we cannot choose the letter we chose from the first group. The third group has 24 items, since we cannot choose the letter we chose from the first two groups. The digits are similar. Thus the number of possibilities is:  $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 = 11,232,000$

### 1.5.2 Permutations

An ordered arrangement of  $r$  distinct items is called a *permutation*. The number of ways of ordering  $r$  items drawn from a group of  $n$  distinct items is designated  $nPr$ .

Before we give the formula for the number of possible permutations, let's look at some examples so we can get an intuitive understanding of what is going on.

**Example.** How many different ordered arrangements are there of the letters  $a$ ,  $b$ , and  $c$ ?

In this case, we can actually write them all out:  $abc$ ,  $acb$ ,  $bac$ ,  $bca$ ,  $cab$ ,  $cba$ .

From this we see that there are 6 possible ordered arrangements. We can also do this using the “choosing-from-groups” approach. For the first letter, we have 3 to choose from; for the second, we have only 2; and for the third, there is only one remaining letter. Thus the number of permutations is:

$$3 \cdot 2 \cdot 2 = 3! = 6$$

Similarly, the number of ordered arrangements of  $n$  distinct symbols is:

$$n \cdot (n - 1) \cdot (n - 2) \cdots 2 \cdot 1 = n!$$

The symbol  $n!$  is the *factorial* operation, and it is defined exactly as written above.

**Example.** Every December (since 1996), the city of Ithaca, NY hosts the International Rutabega Curl<sup>4</sup> (contestants must supply their own rutabega). Gold, silver, and bronze medals are given to the top three finishers. If 100 contentants enter the rutabega curl, how many possibilities are there for the winners?

---

<sup>4</sup><http://www.rutabagacurl.com>

For the gold medalist, we have 100 contestants to choose from. Since you cannot win more than one medal, we choose from 99 contestants for the silver medal and 98 contestants for the bronze medal. The number of medal possibilities is:

$$100 \cdot 99 \cdot 98 = 970,200$$

We can generalize this in the following formula:

The number of permutations (ordered arrangements) of  $r$  items drawn from a group of  $n$  distinct items is:

$${}_nP_r = n(n-1)(n-2) \cdots (n-r+1) = \frac{n!}{(n-r)!}$$

To see this, there are:

- $n$  ways to choose the first item
- $n-1$  ways to choose the second item
- ...
- $n-r+1$  ways to choose the  $r$ th item

Multiply these together to get the permutation formula. To get the “factorial form” of the permutation formula, we have:

$${}_nP_r = n(n-1)(n-2) \cdots (n-r+1) \frac{(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

where  $n! = n(n-1) \cdots 2 \cdot 1$ , and we define  $0! = 1$ .<sup>5</sup>

### 1.5.3 Combinations

Consider the following example, which is modification of the Rutabega Curl example above.

**Example.** 100 students buy raffle tickets. Three names are chosen from a hat uniformly at random to win a free sandwich at Eastside Pockets<sup>6</sup>. How many possibilities are there for the winners?

How is this problem fundamentally different from the Rutabega Curl? Here, the three prizes are identical, as opposed to the three distinct medals in the Rutabega Curl. We can think of this problem as selecting 3 items from a group of 100, but *the order in which we select them does not matter*. Let’s look at this in a few stages:

---

<sup>5</sup>This may seem a little arbitrary, but it is convenient. It also makes some sense that there is exactly one way to arrange no items.

<sup>6</sup>A popular eatery on Thayer St. near Brown University

1. Let's pretend for a moment that the order of selection matters. Then, as in the Rutabega Curl, there are  $100 \cdot 99 \cdot 98$  possibilities.
2. Compared to the the case where order matters, is the number of possibilities greater, fewer, or the same? There must be fewer possibilities since, for example, choosing the numbers  $(1, 2, 3)$  from the hat in that order is the same as choosing them in the order  $(3, 2, 1)$ .
3. How many permutations correspond to choosing the numbers 1, 2, and 3 from the hat? We can write out all the permutations:

$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), \text{ and } (3, 2, 1)$

This gives us a total of 6 permutations. But this number is  $3!$ , the number of ordered arrangements of 3 objects. Does this make sense that this should be the case.

4. Since we have shown that 6 permutations correspond to one possibility of winners, all we need to do is divide the number of permutations by 6. Thus the total number of possibilities for the winners is:

$$\frac{100 \cdot 99 \cdot 98}{6}$$

An unordered arrangement of  $r$  distinct items is called a *combination*. The number of combinations of  $r$  items drawn from a group of  $n$  distinct items is denoted  $\binom{n}{r}$  (or sometimes  $nCr$ ). We can think of this as:

1. The number of subsets of size  $r$  which can be formed from a group of  $n$  distinct objects
2. The number of ways to select  $r$  items from a group of  $n$  distinct items, where order does not matter

Using the logic from the previous example, we can deduce the formula for  $\binom{n}{r}$ . All we have to do is take the number of permutations and divide by  $r!$ , which is the number of ordered arrangements (permutations) of  $r$  items.

The number of combinations (unordered arrangements) of  $r$  items drawn from a group of  $n$  distinct items is:

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

**Example.** The Brown crossword puzzle club consists of 5 undergraduate and 7 graduate students.

1. How many different committees of 2 undergrads and 3 grad students can be formed?

There are  $\binom{5}{2}$  possible groups of 2 undergrads, and  $\binom{7}{3}$  possible groups of 3 grad students. We multiply these together (basic principle of counting) to get:

$$\binom{5}{2} \binom{7}{3} = \frac{5!}{2!3!} \frac{7!}{3!4!} = \frac{5 \cdot 4}{2 \cdot 1} \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 10 \cdot 35 = 350$$

2. What if two of the grads students refuse to be on the same committee?

There are  $\binom{7}{3} = 35$  possible groups of 3 graduate students. How many contain both of the two students who refuse to serve together? To make a three-person committee which includes these two, you need to include the two rival students plus one other student selected from the five remaining graduate students. Thus there are  $\binom{2}{2} \binom{5}{1} = 5$  committees which include the two rival students. Subtracting from 35, there are 30 committees which don't include both rival students. As above, we multiply to get  $10 \cdot 30 = 300$  possible committees.

Computing probabilities of poker hands are classic problems in probability. Since the order of the cards in a poker hand does not matter, these problems involve combinations.

**Example.** You are dealt a five-card poker hand. What is the probability of getting a full house (three cards of one number plus two cards of another number)

There are 52 cards in a poker deck, and a poker hand is 5 cards. Since the order of the cards dealt does not matter, there are  $\binom{52}{5}$  possible five-card poker hands. To count the number of hands which give us a full house, we use the following procedure:

1. Choose the number for the three-of-a-kind. There are  $\binom{13}{1}$  ways to do that. Now select three out of the four cards of this number. There are  $\binom{4}{3}$  ways to do that. Thus there are  $\binom{13}{1} \binom{4}{3}$  ways to choose the three-of-a-kind.
2. Choose the number for the pair. There are  $\binom{12}{1}$  ways to do that, since we have already chosen one of the numbers for the three-of-a-kind. Now select two out of the four cards of this number. There are  $\binom{4}{2}$  ways to do that. Thus there are  $\binom{12}{1} \binom{4}{2}$  ways to choose the pair once the three-of-a-kind has been chosen.
3. Multiply all of these together to get  $\binom{13}{1} \binom{4}{3} \binom{12}{1} \binom{4}{2}$  poker hands which are full houses.

To get the probability of a full house, we divide by the number of poker hands (the size of the sample space) to get:

$$\frac{\binom{13}{1} \binom{4}{3} \binom{12}{1} \binom{4}{2}}{\binom{52}{5}} = 0.00144$$

It is worth noting that the number of full houses is not  $\binom{13}{2} \binom{4}{3} \binom{4}{2}$ , i.e. choose two numbers, then choose a three-of-a-kind, then choose a pair. Why does this not work? This method does not distinguish between, say, 33366 and 33666 (i.e. it treats these as the same), and these are two distinct full houses. Since this undercounts by a factor of two, you can check that if you multiply this by two, you get the correct answer above.

Combinations are incredibly useful. We can use them in some rather surprising cases.

**Example.** Consider the binary string 11000. How many distinct orderings are there of the digits in this string?

At first, this does not appear to involve combinations at all, since we are looking for orderings, and combinations are used where order does not matter. Let's look at this problem in a different way. Consider the five-element set  $A = \{a, b, c, d, e\}$ . There are  $\binom{5}{2}$  two-element subsets of  $A$ . One way to describe subsets of  $A$  is to make a table. The columns of the table represent the elements of  $A$ :  $a, b, c, d$ , and  $e$ . Each row corresponds to a subset of  $A$ : a 1 indicates that the element is in the subset, and a 0 indicates that it is not. Here is an example table, where we depict two two-element subsets.

$a$	$b$	$c$	$d$	$e$	subset
0	0	1	1	0	$\{c, d\}$
1	0	1	0	0	$\{a, c\}$

Notice that two-element subsets match up exactly with orderings of 11000! Thus we see that there are  $\binom{5}{2}$  rearrangements of the binary string 11000.

In general, if you have a string of length  $n$  composed of two symbols, and you have  $r$  of one symbol (thus  $n - r$  of the other), the number of distinct orderings of the string is  $\binom{n}{r}$ . Equipped with this, let's return to the antenna example from the beginning of the section.

**Example.** A communication system consists of  $n$  antennas arranged in a line. Exactly  $m$  out of the  $n$  antennas are defective. The system is functional if no two consecutive antennas are defective. Assuming that each linear arrangement of the antennas is equally likely, what is the probability that the system will be functional?

First we need to figure out how many possible arrangements there are, i.e. the size of our sample space. If we use the digit 1 to represent a functional antenna, and the digit 0 to represent a defective antenna, then each arrangement can be represented by a binary string of length  $n$  consisting of  $m$  0s and  $n - m$  1s. Thus from what we learned above, there are  $\binom{n}{m}$  possible arrangements.

How many of these arrangements are functional? Let's draw a picture! Imagine we have the  $n - m$  functional antennas in a line. We will represent each functional antenna by the symbol \*, and the spaces between the functional antennas by a vertical bar |. We will also place a space to the far right and to the far left:

$$| * | * | * | * | * | * | \dots | * | * |$$

For the system to be functional, each space | can contain at most one defective antenna. There are  $n - m + 1$  spaces to choose from, and  $m$  defective antennas to place. Thus there are  $\binom{n-m+1}{m}$  functional arrangements. Dividing this by the total number of arrangements

$\binom{n}{m}$ , the probability that the system is functional is:

$$\frac{\binom{n-m+1}{m}}{\binom{n}{m}}$$

For a sanity check, let's calculate this for the case where  $n = 4$  and  $m = 2$ , which we did manually above:

$$\frac{\binom{n-m+1}{m}}{\binom{n}{m}} = \frac{\binom{3}{2}}{\binom{4}{2}} = \frac{3}{6} = \frac{1}{2}$$

This agrees with what we found above!

**Example.** I buy an assortment of 10 bagels from Bagel Gourmet on Thayer St.<sup>7</sup>. There are five types of bagels to choose from: everything, onion, poppy, raisin, and sesame. How many different assortments of bagels can I bring home? Assume there are at least 10 of each kind of bagel in the store.

We can model the problem in a way that lets us use combinations. Imagine you buy the bagels in a special box. The bagels are arranged in a line, and are always in alphabetical order: everything, onion, poppy, raisin, sesame. There are four dividers, one between each bagel type, and if a bagel type is not present, we just put the dividers next to each other. Here is an example of three bagel boxes with their dividers. Bagels are designated by \* and dividers by |.

** * **** ** **	2 everything, 1 onion, 3 poppy, 2 raisin, 2 sesame
***  ****  **	3 everything, 3 poppy, 2 sesame
*****	10 everything

From the picture, we see that an assortment of 10 bagels can be represented as a linear string of 14 symbols: 10 \* representing the bagels and 4 | representing the dividers. The number of assortments is the same as the number of distinct orderings of this linear string:  $\binom{14}{4} = 1001$

---

<sup>7</sup>Having lived in Brooklyn for many years, I am only *slightly* obsessed with bagels.