

# Basic probability and statistics

Philip Schulz  
Christian Schaffner

last modified: September 11, 2015

## Contributors

While we strive to continuously update this script and keep it on an acceptable level of grammaticality and mathematical correctness, it is unavoidable that some mistakes creep in. We are therefore utterly grateful to our contributors who have helped improving the script and would like to acknowledge their contributions here.

- Philip Michgelsen has corrected a mistake in the definition of event spaces in chapter 1.

# Chapter 1

## Basic Probability And Combinatorics

### Notational conventions

In this script we make use of certain notational conventions. We **bold-face** newly introduced technical terms on first mention. Those are the terms whose definitions you are expected to know by heart in this and following courses. *Italics* serve the purpose of highlighting passages in the script but also to discriminate linguistic examples from the rest of the text. Occasionally, we will point to online references outside of this script. The corresponding links are coloured in [blue](#) and you are encouraged to click them.

We denote sets with uppercase letters and overload notation by using  $|\cdot|$  as both a function that yields the cardinality of a set and the length of a sequence. Besides using standard notation for set union and intersection we denote the complement of a set  $S$  with respect to another set  $X$  by  $S \setminus X$ .

### 1.1 Introduction

#### 1.1.1 Why study probability theory?

The fact that you have picked up this script and started reading it demonstrates that you already have some interest in learning about probability theory. This probably means that you also have some conception of what probability theory is and what to do with it. Nevertheless, we will take the opportunity to quickly give you some additional motivations for studying probability theory.

This script is all about formalizing the notion of probability. In particular, we are interested in giving a formal interpretation to statements like “A is more probable than B”. Let us take a simple example to demonstrate

why this is useful: Suppose it is Monday and you have a date scheduled for Friday. Obviously you want to impress your date. Unluckily, however, you have tendency to be broke come weekends. The decision you have to make now is whether to take your date to a fancy restaurant (the impressive but expensive option) or to just go for drinks (the cheaper option). On what basis can you make this decision? Well, you can ask yourself whether it is more likely that you are broke on Friday night or not. If you think that you being broke is more probable than you going for drinks, otherwise you opt for the fancy restaurant.

The above is an example where we have used the intuitive notion of probability to assist us in decision making. The first part, the computation of the probabilities of events (e.g. you being broke or not) is something that we are going to develop in some detail in this script. The second part, the development of a so-called *decision rule* (e.g. to plan for the circumstances that are most probable to occur in the future) is something that will be covered in later courses.

Here is a second example of what one can do with probability theory. Assume you want to invest in the stock market. You will be putting in some money now and then you want to cash in on your gains (or losses) in ten years time, say. Notice that this time around simply asking whether it is more probable that your stock has risen or fallen in price is not enough. Even if your stock is worth more in ten years than it was when you bought it, the absolute increase may be so miniscule that you could have found much better investment options that would have yielded more gains. Worse even, if your gain is a smaller percentage of your original capital than the overall inflation that occurred during the ten years of your investment, you will actually have incurred a loss in terms of pure market power! So instead of asking whether or not your stock will be worth more than what it was when you first bought it, you should rather ask how much of an absolute gain you can expect from your investment. This second application of probability theory, the computation of expectations over real values, is something we are going to cover in this script, as well.

Alright, we hope that this has gotten you excited for the rest of the script. Let's get going!

## 1.2 Sample spaces and events

The whole of probability theory is based on assigning probability values to elements of a **sample space**. The members of the sample space are referred to as **outcomes** or **samples**.

**Definition 1.1 (Sample Space)** A sample space is any *Borel set*  $\Omega$ . We denote the members of a sample space by  $\omega \in \Omega$ .

Standard examples of sample spaces are the flipping of a coin and the rolling of a die. Formally, the sample space of a die roll is  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . The sample space of a coin toss would consist of heads and tails. However, it is often more convenient to represent outcomes numerically. In the context of this course, we will achieve this by imposing any total order on the sample space and then identifying the outcomes with the positions they occupy in the corresponding ordered list. In this spirit we let the sample space of a coin toss be  $\Omega = \{1, 2\}$  where 1 represents heads and 2 represents tails, say (the other way around would be just as fine).

More generally, we denote a sample space with  $n$  members as  $\Omega = \{1, \dots, n\}$ . A useful metaphor that we will often use is to think of generating an outcome from a sample space as a blind draw from an urn with  $n$  balls that are numbered and possibly coloured but otherwise indistinguishable. The rolling of a die, for example, corresponds to drawing a ball from an urn with balls numbered 1 to 6. A somewhat more involved example is that of writing an English sentence of six words, for example the sentence: *To be or not to be*. The process of writing this sentence can be conceptualized as drawing six balls from an urn that contains balls corresponding to words in the English language<sup>1</sup>. Note that this will be a rather large urn as [the vocabulary of the English language has already exceeded 1 million words](#).

In our sample spaces as defined above, it is easy to distinguish individual outcomes. However, often times we do not care about the outcomes themselves but about properties that some of them share. In the die example we might be only interested in whether the outcome is even or odd. Transferring this scenario to the urn metaphor we would colour the balls with odd numbers green and the balls with even numbers red. Again, any other colours are just as fine. All that matters is that we can discriminate a member of  $E = \{2, 4, 6\}$  from a member of  $O = \{1, 3, 5\}$ . We do *not* need to discriminate between the outcomes that are members of the same set! In this particular setting  $E$  and  $O$  are the **events** that we are interested in.

**Definition 1.2 (Event)** *An event  $A$  is any subset  $A \subseteq \Omega$ .*

Events are what usually interests us in probability theory. Just as with outcomes, we can also define the notion of an event space.

---

<sup>1</sup>This is obviously a very unrealistic conception of how English sentences are written as it totally ignores the fact that the words in a sentence are dependent on each other and have to be placed in a particular order.

**Definition 1.3 (Event space)** An event space associated with a sample space  $\Omega$  is a set  $\mathcal{A}$  such that

1.  $\mathcal{A}$  is non-empty
2. If  $A \in \mathcal{A}$  then  $A \subseteq \Omega$
3. If  $A \in \mathcal{A}$  then  $\Omega \setminus A \in \mathcal{A}$
4. If  $A, B \in \mathcal{A}$  then  $A \cup B \in \mathcal{A}$

Notice that since  $\emptyset \subseteq S$  for any set  $S$  we always have  $\Omega \in \mathcal{A}$  by item 3.

**Exercise 1.1** You can also arrive at the conclusion that  $\Omega \in \mathcal{A}$  always holds in a different (and arguably more cumbersome) way. How so?

The fact that event spaces are closed under the set complement operation is very convenient. Say I organized a dinner party and invited 10 people. The day after you ask me if more than 8 people actually showed up. I just answer that I was very disappointed that my friends Mary and Paul did not come. Although I did not directly address your question you know that the answer is negative. After all, I informed you that the complement event of the event you asked about had occurred.

**Exercise 1.2** In the above party example, what is the sample space? What is the smallest possible event space that is necessary to model the situation just described?

In general, we will not worry too much about constructing an event space every time we encounter a new problem. The **power set** of the sample space conveniently happens to fulfil all the requirements we have for event spaces, so we will just always use it. Thus, all we will ever need to worry about is the construction of sample spaces since we now know how to construct event spaces from them in a simple manner. In case you are a bit rusty, here is a reminder of what a power set is.

**Definition 1.4 (Power Set)** The power set  $\mathcal{P}(S)$  of any set  $S$  is defined as  $\mathcal{P}(S) := \bigcup_{s \subseteq S} s$ .

In general, this leaves us with the pair  $(\Omega, \mathcal{P}(\Omega))$ . For outcomes in a sample space, let us stress again an important difference, namely that  $\omega \in \Omega$  but  $\{\omega\} \in \mathcal{A}$ .

### 1.3 Some basic combinatorics

Combinatorics is the mathematics of counting. Counting is of course a very basic problem that may be solved by just looking at each element of a

set. However, this naïve procedure is often unreasonably time consuming. Moreover, it does not allow us to make general statements about sets of any size, i.e. sets of size  $n$ .

In order to assess the size of our sample spaces, we would like to make such general statements. The reason is that when we are dealing with probability we often start from **uniform probabilities** on the sample space where by uniform probability we simply mean the value  $\frac{1}{|\Omega|}$ . This is the probability we will assign to each and every  $\omega \in \Omega$ . We now say that all the elements in our sample space are equally probable. Note that at this point we are using probabilities solely for the purpose of motivating combinatorics which is kind of a hack because we haven't even told you yet what a probability is. However, we hope that you find the idea of uniform probabilities somewhat intuitive.

Let us start from scratch: What is the cardinality (size) of the sample space of a die roll? It is 6 because  $|\{1, 2, 3, 4, 5, 6\}| = 6$ . Now what if we roll two dice? The sample space for each individual die is already known. Let us call it  $\Omega_1$ . The sample space for the rolling of two dice is then just the Cartesian product of two such sample spaces, i.e.  $\Omega_2 = \Omega_1 \times \Omega_1 = \{(x, y) | x \in \Omega_1, y \in \Omega_1\}$ . Since the cardinality of the Cartesian product of two sets  $S$  and  $S'$  is  $|S| \times |S'|$  we conclude that  $|\Omega_2| = |\Omega_1 \times \Omega_1| = |\Omega_1| \times |\Omega_1| = |\Omega_1|^2 = 36$ .

Unsurprisingly, this method of performing a draw from the same sample space (urn) multiple times generalizes to any number of times  $n > 2$ . Nicely enough, it also generalizes to sets of different sizes (again by the Cartesian product argument from above). However, we have to impose one important restriction on the use of this technique: it may only be applied when the sample spaces are independent, i.e. when the outcome of one space does not affect the outcome of the other. Often times, we will simply assume that this is the case, though.

The technique of inferring the size of a complex sample space from the sizes of the sample spaces it is constructed from is known as the **basic principle of counting**.

**Definition 1.5 (Basic principle of counting)** *The basic principle of counting states that if two draws from sample spaces of size  $M$  and  $N$  respectively are performed independently of each other then the sample space composed from them has size  $M \times N$ .*

**Exercise 1.3** *Let us assume that a football game is played for strictly 90 minutes. Both teams start with 11 players. A red card to a player results in that player being sent off the pitch. According to the rules of football, the game is stopped prematurely when either team has only 6 or fewer players remaining on the pitch. We are now interested in how many possible situations (we assume that situations occur in one-minute*

intervals) there are in which the game still progresses, one or more red cards have been issued and exactly four goals have been scored. Give the corresponding sample space and its size.

Note that up to now we have implicitly assumed that we would put every drawn ball back into the urn. This is also referred to as **sampling with replacement**. Let us now look at problems for **sampling without replacement**, i.e. problems where we are shrinking our sample space at each draw. One class of such problems is known as **permutation** problems.

**Definition 1.6 (Permutation)** A permutation on a set  $S$  is a bijection  $\sigma : S \rightarrow S : s \mapsto \sigma(s)$ .

Often times people also use the word permutation to refer to the image of a set under a permutation. What we need permutations for in practice is the reordering of ordered sets (which we will call lists). For example the permutations of the list  $L = (1, 2, 3)$  are:

- |  |                           |
|--|---------------------------|
| • $\sigma_1 = \{1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 3\}$ | $\sigma_1(L) = (1, 2, 3)$ |
| • $\sigma_2 = \{1 \mapsto 1, 2 \mapsto 3, 3 \mapsto 2\}$ | $\sigma_2(L) = (1, 3, 2)$ |
| • $\sigma_3 = \{1 \mapsto 2, 2 \mapsto 1, 3 \mapsto 3\}$ | $\sigma_3(L) = (2, 1, 3)$ |
| • $\sigma_4 = \{1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 1\}$ | $\sigma_4(L) = (2, 3, 1)$ |
| • $\sigma_5 = \{1 \mapsto 3, 2 \mapsto 1, 3 \mapsto 2\}$ | $\sigma_5(L) = (3, 1, 2)$ |
| • $\sigma_6 = \{1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1\}$ | $\sigma_6(L) = (3, 2, 1)$ |

The way to think about a permutation as a draw from an urn is to look at each of the positions in the list in turn and insert an element from  $S$ . Since a permutation is a bijection, we can only use each  $s \in S$  exactly once. This is precisely what it means to sample without replacement. Once a ball is drawn, it is removed from the urn. Let us make this effect concrete in the above example. For position one we have three elements to choose from. Hence we are dealing with a sample space of size 3. Position two still leaves us 2 choices, giving us a sample space of size 2. Finally, the element in the last position is totally determined as we are dealing with a sample space of size 1.

Applying the basic principle of counting we now know that there are  $3 \times 2 \times 1$  permutations of the list  $(1, 2, 3)$ . Incidentally, this proves our above example to be correct. More generally, if we have to reorder a list with  $n$  distinct elements (or draw without replacement from an urn with  $n$  numbered balls), there are  $n \times (n - 1) \times \dots \times 2 \times 1$  permutations. Since this is pretty painful to write down we introduce a more succinct notation, provided by the **factorial** function.



**Definition 1.7 (Factorial)** *The factorial  $n!$  of a non-negative natural number  $n \in \mathbb{N}$  is defined recursively as*

- $0! = 1$
- $k! = k \times (k - 1)!$  for  $0 < k \leq n$

From the above discussion we can now conclude that the number of permutations on a set or list of size  $n$  is  $n!$ .

We can also define the notion of a  $k$ -permutation on a set  $S$  of size  $n$  such that  $k < n$ . This means we are still drawing without replacement but we do not fully empty the urn. The reasoning for how many of those  $k$ -permutations there are remains exactly the same. There are  $n \times (n - 1) \times (n - k + 2) \times (n - k + 1)$  such permutations (make sure you understand why!). In order to ease notation we can again sneak in the factorial through multiplying this number with 1 in disguise. Concretely, we write

$$\begin{aligned} & n \times (n - 1) \times \dots \times (n - k + 2) \times (n - k + 1) \times 1 \\ &= n \times (n - 1) \times \dots \times (n - k + 2) \times (n - k + 1) \times \frac{(n - k)!}{(n - k)!} \\ &= \frac{n!}{(n - k)!} \end{aligned}$$

for the number of  $k$ -permutations on a set of size  $n$ .

We will not see  $k$ -permutations all that often in this script but they constitute a helpful stepping stone to another concept that will be of crucial importance. Let us draw  $k$  balls from an urn with  $n$  balls where  $k \leq n$  and disregard the order in which we draw them. A classical example of such a setting would be the lottery where you are only interested in the balls drawn but not in the order in which they were drawn. We already know that for a set of  $k$  balls there are  $\frac{n!}{(n - k)!}$  orders in which we can draw them, as this is a  $k$ -permutation on our urn. Now, though, we need to get rid off the different orderings. This is to say that we want to count each set of  $k$  balls that we can draw only once and not once per permutation of it. Luckily, we know how many permutations of a set of size  $k$  there are, namely  $k!$ . Thus we divide out this number of permutations, yielding  $\frac{n!}{(n - k)! \times k!}$  as the number of possible ways to draw  $k$  *different* balls from an urn with  $n$  balls. At this point we should take a break and pat our own backs. After all, we have just derived one of the most important combinatorial formulas, which is known as the **binomial coefficient**.

**Definition 1.8 (Binomial co-efficient)** *The binomial coefficient  $\binom{n}{k}$  is defined as*

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

for  $0 < n, 0 \leq k \leq n$ . It counts the number of ways to sample  $k$  distinct elements from a set with a total of  $n$  elements without regard to the order in which they are drawn. For this reason, it is pronounced “ $n$  choose  $k$ ”.

**Exercise 1.4** *In the German lottery you have to bet on a set of 6 numbered balls to be drawn out of a total of 49 balls. Assuming that each ball is equally likely to be drawn, what is the chance of an individual bet to win the jackpot? The Dutch lottery is slightly more involved. They also draw an additional coloured ball from 6 coloured balls. In order to win the jackpot you need to have the number-colour combination right. What is your chance here?*

The binomial coefficient will become crucially important later on. A common application, that you will see in this and other courses is counting the number of bit strings with certain properties. A bit is a variable that can take on values in  $\{0, 1\}$ . By the basic principle of counting there are  $2^n$  bit strings of length  $n$ . How many bit strings of length 5 are there that contain exactly 3 ones? Well, there are  $2^5 = 32$  bit strings of that length in total and  $\binom{5}{3} = 10$  of them contain exactly three ones. Unsurprisingly, this is the same number of 5-bit strings with exactly 2 zeros. The moral lesson here is that  $\binom{n}{k} = \binom{n}{n-k}$  as can be easily seen from the definition. Some other trivia about the binomial coefficient are that  $\binom{n}{0} = \binom{n}{n} = 1$ . Again, this follows directly from the definition. Somewhat trickier is the fact that  $\binom{n}{1} = \binom{n}{n-1} = n$ . Can you derive this?

We can straightforwardly generalize the idea of the binomial coefficient to choosing more than just one set of objects. This means that instead of just looking at red versus non-red balls, say, we now distinguish between all the colours in our urn. For our strings this means that we move away from bit strings to strings with large alphabets, e.g. strings written in the English alphabet (which has 26 letters). Let's say we have  $r$  red,  $b$  blue,  $g$  green and  $y$  yellow balls in our urn such that  $n = r + b + g + y$  is the total number of balls in the urn. How many different colour sequences can we draw? Well, we first arrange the  $r$  red balls in  $r$  out of  $n$  positions. This can be done in  $\binom{n}{r}$  ways. We then place the  $b$  blue balls in  $\binom{n-r}{b}$  ways. Next, we place the  $g$  green balls in  $\binom{n-r-b}{g}$  ways. Finally, we place the remaining yellow balls deterministically in the remaining positions since  $\binom{n-r-b-g}{y} = \binom{y}{y} = 1$ . We compute the total number of arrangements as

$$\begin{aligned}
(1.1) \quad & \binom{n}{r} \binom{n-r}{b} \binom{n-r-b}{g} \binom{n-r-b-g}{y} = \\
(1.2) \quad & \frac{n!}{r! \times (n-r)!} \times \frac{(n-r)!}{b! \times (n-r-b)!} \times \frac{(n-r-b)!}{g! \times (n-r-b-g)!} \times 1 = \\
(1.3) \quad & \frac{n!}{r!b!g!y!}
\end{aligned}$$

Observe that the last equality follows because many of the factorials cancel and because we know that  $n - r - b - g = y$ . We have now worked with only four colours, but the general case follows directly by induction on the number of colours (with the binomial coefficient as base case). Thus, we can define the **multinomial coefficient**.

**Definition 1.9 (Multinomial co-efficient)** *The multinomial coefficient for choosing  $k$  sets of objects with size  $m_k$  from a total of  $0 < n = \sum_{i=1}^k m_i$  objects is*

$$\frac{n!}{\prod_{i=1}^k m_i!}$$

## Further material

For a slow and thorough introduction to combinatorics, see [Faticoni \(2013\): Combinatorics](#). At the ILLC, there is [a biannual course on combinatorics](#), taught by Ronald de Wolf. Online, Princeton also offers [a course on combinatorics](#).