

# Probability Theory and Statistics

Malcolm Connolly

Semester 2, 2023



# Contents

<b>1</b>	<b>Introduction to Probability</b>	<b>5</b>
1.1	Frequentist perspective . . . . .	5
1.2	Naive probability . . . . .	7
1.3	Complements and mutual exclusivity . . . . .	9
1.4	Outcomes and counting . . . . .	11



# Chapter 1

## Introduction to Probability

Some things that happen are entirely predictable. For example, if one drops a ball from a height, we know it will hit the ground. Things that happen like this can be described as *deterministic*. You may have heard people talk about things being written in the stars, or their fate, or destiny. The opinion that all things are pre-determined is called *determinism*.

However, even if one is a determinist, one will have to live with uncertainty. In our everyday lives we can think of examples where things happen that we cannot predict; a bus may be late, it may rain, or one might win the lottery. To one living with uncertainty, it is reasonable to quantify this uncertainty and act assuming outcomes are not pre-determined. If the outcome is not pre-determined then it is called *random*.

The Mathematics of random phenomena is called Probability Theory. Most people have an intuitive idea of what is meant by probability or chance. Unfortunately Probability Theory is a subject in which there are endless examples of seemingly simple questions that turn out to be very complicated or have severely counter-intuitive answers.

### 1.1 Frequentist perspective

We need to start with some terminology.

**Definition 1.1.** An *experiment* is any procedure which happens at random with at least two different outcomes. For example rolling a die and observing the score is a statistical experiment. If the experiment is repeatable then each repetition is called a *run*.

By calculating the number of times an event occurs divided by the number of runs one can estimate the theoretical probability. The idea is that the relative

cumulative frequency of outcomes will tend to the actual probability in the long run. This perspective of probability is called *Frequentist*, and is incredibly useful in practice.

gganim\_plot0100.png

We will recreate a plot like this in labs.

**Example 1.1.** Suppose we toss a 10 coins 10 times and the results are recorded in the table below, draw the graph of relative frequency.

Run	1	2	3	4	5	6	7	8	9	10
Outcome	6H	3H	3H	1H	6H	3H	6H	5H	5H	7H

The cumulative relative frequencies are calculated as the cumulative number of flips divided by the cumulative number of heads:

Cumulative flips $n$	10	20	30	40	50	60	70	80	90	100
Cumulative heads $a_n$	6	9	12	13	19	22	28	33	38	45
Relative Frequency	0.6	0.45	0.4	0.325	0.38	0.367	0.4	0.413	0.422	0.45

In this course we will learn some R programming. R is a free open-source software language suitable for doing many probability and statistical calculations. The following R code will make a list of two outcomes Heads or Tails and create a sample of 10 random outcomes.

```
outcomes <- c("Heads","Tails")
sample(outcomes, 10, replace=TRUE)
```

```
## [1] "Tails" "Tails" "Heads" "Heads" "Tails" "Tails" "Heads" "Tails" "Heads"
## [10] "Heads"
```

**Definition 1.2.** If a statistical experiment has  $n$  runs, and the outcome  $A$  happens a cumulative number of times depending on  $n$  which we can call  $a_n$ , then the *frequentist probability* of the outcome  $A$ , written  $P(A)$ , is the limit:

$$P(A) = \lim_{n \rightarrow \infty} \frac{a_n}{n}$$

So if it is possible to repeatedly run an experiment, frequentist methods are very useful for finding an approximation of the true theoretical probability.

Not all is so simple, consider the following questions. What is the probability that there is life on other planets? What is the probability that the Conservatives win the next general election?

These events are not like flipping a coin, and so it is not possible to find a frequentist interpretation for their probability.

## 1.2 Naive probability

We may not have the time or resources to do many thousands of runs. Therefore we also need to be able to evaluate the theoretical probability directly and exactly.

**Definition 1.3.** The *sample space* is a set whose elements are outcomes of an experiment. The sample space is denoted by the greek letter  $\Omega$ .

**Example 1.2.** If we pick a person at random on the street and ask them the month of their birthday, we can let

$$\Omega = \{\text{Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec}\}.$$

**Definition 1.4.** An *event* is a subset of the sample space  $\Omega$ .

**Example 1.3.** As in example 1.2, let  $L$  be the *event* that the month is a long month (i.e. has 31 days). Then

$$L = \{\text{Jan, Mar, May, Jul, Aug, Oct, Dec}\}.$$

Let  $R$  be the *event* that there is a letter  $r$  in the name of the month when written fully. Here,

$$R = \{\text{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec}\}$$

**Definition 1.5.** Naively the the probability of an event  $A$  should be the number of elements of the set  $A$  divided by the size of the sample space  $\Omega$ . That is,

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

In our example 1.3 above:

$$P(R) = \frac{|R|}{|\Omega|} = \frac{8}{12} = \frac{2}{3},$$

and,

$$P(L) = \frac{|L|}{|\Omega|} = \frac{7}{12}.$$

**Example 1.4** (Coin Tossing). Toss a fair coin twice and record the possible outcomes. Let

$$A = \{\text{exactly one coin is Heads}\}$$

and

$$B = \{\text{neither coin is Heads}\}$$

The sample space here is  $\Omega = \{HH, HT, TH, HH\}$ .

Events  $A$  and  $B$  correspond to:

$$A = \{HT, TH\}$$

and

$$B = \{TT\}$$

Hence  $P(A) = \frac{2}{4} = \frac{1}{2}$ , and  $P(B) = \frac{1}{4}$ .

**Example 1.5** (Two dice). Two dice are thrown, what is the probability that the total number of dots is:

- a) equal to 7
- b) equal to 3
- c) greater than 5
- d) an even number

*solution*

The sample space here is  $\Omega = \{(n_1, n_2) : n_1, n_2 \in \{1, 2, 3, 4, 5, 6\}\}$ . However, not all sums are equally likely, which is best seen in a table.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- a)  $\frac{6}{36}$
- b)  $\frac{2}{36}$
- c)  $\frac{26}{36}$
- d)  $\frac{18}{36}$

For infinite sets there is a problem with the naive definition 1.5. Consider the following:

**Example 1.6.** Suppose a random unit vector is rotated about the origin anti-clockwise, making an angle  $\theta$  with the positive  $x$ -axis. What is the probability that this angle is acute?

There are a continuum of infinitely many such angles. The naive definition says  $\frac{\infty}{\infty}$ , which is absurd.

Intuitively, the answer *should* be  $\frac{1}{4}$ .



### 1.3 Complements and mutual exclusivity

In any case, as events are subsets of the sample space  $\Omega$  and follow the rules of set theory, and so it is important to know some set notation, definitions and results. Below is a recap of the important definitions.

**Definition 1.6.** The *union* of  $A$  and  $B$  is written:

$$A \cup B = \{x \in \Omega : x \in A \text{ or } x \in B\}.$$

In Mathematics or is inclusive, which means we do not need to say “or both” as this is included in the union.

**Definition 1.7.** The *intersection* of  $A$  and  $B$  is written:

$$A \cap B = \{x \in \Omega : x \in A \text{ and } x \in B\}.$$

**Definition 1.8.** The empty set  $\emptyset$  is the set of no elements. As sets  $A$  and  $B$  are called disjoint if they have no elements in common, that is,

$$A \cap B = \emptyset.$$

In Probability Theory disjoint events are called *mutually exclusive*.

**Definition 1.9.** The *complement* of an event  $A$  is the event  $A^c = \{x \in \Omega : x \notin A\}$ . Note  $A \cap A^c = \emptyset$ . In words this means: any event is mutually exclusive with its complement.

**Example 1.7.** Suppose the event is throwing a die. The event is that one throws an even number. The complement is that one throws an odd number.

**Example 1.8.** Suppose the event is that a random student has no siblings. The complement is not that they have one sibling. The complement is that they have *at least* one sibling.

A theorem which we will not prove is De Morgan’s laws

**Theorem 1.1 (DE MORGAN’S LAWS).** *The complement of a union is the intersection of the complements:*

$$(A \cup B)^c = A^c \cap B^c$$

*The complement of an intersection is the union of the complements:*

$$(A \cap B)^c = A^c \cup B^c$$

In this way  $P$  is a ‘measure’ function which maps the subsets of the sample space to the interval  $[0, 1]$ .

**Definition 1.10. *Probability*** is a function whose input is a subset of the sample space  $A \subseteq \Omega$  and whose range is the interval  $[0, 1]$ , such that the following two axioms hold:

- (i) The probability of the whole set of possible events is unity. In the notation:  $P(\Omega) = 1$ .
- (ii) (*additivity*) For any collection of disjoint events  $A_1, A_2, A_3, \dots$  the probability of the union is the sum of the probabilities. In the notation this can be written as

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

The above definition 1.10 is due to the Russian Mathematician Kolmogorov. These axioms help make sense of the infinite case.

Using this definition we can prove the following important results.

**Proposition 1.1 (THE PROBABILITY OF A COMPLEMENT).** *For any event  $A$  we have:*

$$P(A^c) = 1 - P(A).$$

*Proof.* Write  $\Omega = A \cup A^c$ , which is a disjoint union. Then by additivity,

$$P(\Omega) = P(A) + P(A^c)$$

Now by axiom (i) the LHS is 1. □

**Theorem 1.2 (THE PROBABILITY OF A UNION).** *Given any two events  $A$  and  $B$  we have:*

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

*Proof.* The idea is to write  $A$  as a disjoint union of the part that has intersection with  $B$ , and that which does not:  $A = (A \cap B) \cup (A \cap B^c)$ . Hence,

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

If we split  $A \cup B$  in the same way, we obtain  $(A \cup B) \cap B$  and  $(A \cup B) \cap B^c$ . The former is simply  $B$ , and the latter is  $A \cap B^c$ . Again by additivity,

$$P(A \cup B) = P(B) + P(A \cap B^c).$$

Eliminating  $P(A \cap B^c)$  from the two equations above proves the rule. □

We will not be proving all Theorems in this course, neither will I ask you to recount a proof in an exam. You will however have to know how to use these results in applied problems.

**Example 1.9** (Multiple Choice). Suppose a multiple choice test consists of three questions each of which has two options, the correct answer (C) or the wrong answer (W). What is the probability that a student who always randomly guesses the answers gets at least one correct?

$$P(\text{at least one correct}) = 1 - P(\text{all wrong}) \quad (1.1)$$

$$= 1 - \frac{1}{8} \quad (1.2)$$

$$= \frac{7}{8} \quad (1.3)$$

**Example 1.10** (Mode of travel). The table shows the type of journey undertaken by a sample of commuters classified by where they live.

	Town		Rural
Car	40	30	70
Bus	25	5	30
	65	35	100

If an individual is selected at random from this group, find the probability that, they travel by car or live in the town

*solution*

$$P(\text{Car} \cup \text{Town}) = \frac{25+40+30}{100} = 0.95$$

$$P(\text{Car}) + P(\text{Town}) - P(\text{Car} \cap \text{Town}) = \frac{65}{100} + \frac{70}{100} - \frac{40}{100} = 0.95$$

**Example 1.11.** In a particular city 60% of people watch the news in the morning, 50% of people watch the news in the evening and 30% watch both. What is the probability that an individual selected at random watches either the morning news or the evening news.

*solution*

$$P(M \cup E) = 0.6 + 0.5 - 0.3 = 0.8$$

## 1.4 Outcomes and counting

One might imagine that the finite situation is then very simple, and even then we have seen this is not the full picture. One simply counts how many ways an event can happen out of the total number of configurations. This can actually be quite complicated. We will learn some formulae to enable us to count them.

### 1.4.1 Factorials

**Example 1.12** (Three people in a line). In how many ways can three people  $A$ ,  $B$  and  $C$  stand in a line?

*solution*

$ABC, ACB, BAC, BCA, CAB, CBA$  there are 6.

**Definition 1.11.** For any non-negative integer,  $n$  say, we define the factorial of  $n$ , written  $n!$  to be equal to the product of  $n$  and all the numbers less than  $n$  down to 1. That is,

$$n! = n \times (n - 1) \times (n - 2) \times \dots \times 3 \times 2 \times 1$$

**Definition 1.12** (Multiplication Rule). If there are  $n$  ways for some operation to happen, and  $m$  ways for something else to happen, then the total number of ways for the sequence to occur is  $n \times m$ .

**Example 1.13.** MMU assigns each student an 8 digit ID number. How many possible ID numbers are there?

*solution* The first digit is not zero, there are 9 digits from which to choose. All the other digits have 10 choices 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Total =  $9 \times 10^7$ .

**Example 1.14** (objects in a line). The number of ways of arranging  $n$  distinct objects in a line is  $n!$ . This is because there are  $n$  choices for the first number in line, then one fewer choice ( $n - 1$ ) for the second, and so on, until the last one in the line there is only one choice remaining.

**Definition 1.13** (rule of division). The number of ways of arranging  $n$  objects in a line where  $p$  are the same is  $\frac{n!}{p!}$ .

**Example 1.15.**

- a) Suppose you have the letters  $A, A, A, B$  - how many 'words' can be made?
- b) Suppose you have the letters  $A, A, A, B, B$  - how many 'words' can be made?

*solution* a) AAAB, AABA, ABAA, BAAA

There are 4. How to find this number without having to write them down?

You might think  $4!$  but this is thinking each A is different, and so overcounts the same word. By what factor does it overcount? Take one of the words such as ABAA and number each A, one finds rearrangements of 1,2,3:

$A_1BA_2A_3, A_1BA_3A_2, A_2BA_1A_3, A_2BA_3A_1, A_3BA_1A_2, A_3BA_2A_1$ .

The upshot is that you need to divide by the factorial of number of letters that are the same, here  $\frac{4!}{3!} = 4$ .

- b) Here there are 3 of the same letter  $A$ , and 2 of the same letter  $B$ . The correct number is

$$\frac{5!}{3! \times 2!} = 10$$

The words are AAABB, AABBA, ABBA, BBAAA, BABAA, ABABA, AABAB, BAABA, ABAAB, BAAAB. (Here I can systematically list them by considering the number of A's between the B's).

**Definition 1.14** (rule of sum). Given two disjoint events  $A$  and  $B$ , then the size of the union is the sum of the sizes of  $A$  and  $B$ . That is,

$$|A \cup B| = |A| + |B|$$

**Example 1.16.** How many possible MMU IDs start with a 1 or a 3?

*solution*

The IDs are all of the form 1\*\*\*\*\* or 3\*\*\*\*\*. There is only 1 choice for the first digit and  $10^7$  choices for the next digits in either case.

The total number starting with a  $1 \times 10^7 + 1 \times 10^7 = 2 \times 10^7$ .

## 1.4.2 Permutations

**Example 1.17.** Consider the number of ways of placing three of the letters  $A, B, C, D, E, F, G$  in three empty spaces. The first space can be filled in 7 ways, the second in 6 ways and the last in 5 ways.

In total this is  $7 \times 6 \times 5 = 120$

This number can be written as

$$\frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1} = \frac{7!}{(7-3)!}$$

**Definition 1.15** (Permutations). The number of ways of choosing  $k$  distinct items from  $n$  when the order is relevant is

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

Any way of choosing  $k$  distinct items from  $n$  when order matters is called a *permutation*.

**Example 1.18.** My PIN has 4 different digits. How many different such PINs are there?

*solution*

Order matters here - the guess 1234 is different from 4321, for example.

$${}^{10}P_4 = \frac{10!}{(10-4)!} = \frac{10 \times 9 \times \dots \times 2 \times 1}{6!} = 10 \times 9 \times 8 \times 7 = 5040$$

The expression  $10 \times 9 \times 8 \times 7$  can be interpreted as saying there are 10 choices for the first digit, 9 for the second, and so on.

**Example 1.19** (The Birthday Problem). Suppose there are  $k$  people in a room. What is the probability that at least one has the same birthday as someone else in the room?

*solution*

$P(\text{at least one birthday the same}) = 1 - P(\text{all birthdays different})$

The first person could be born on any day there are 365 such days, the second person has to have a different birthday so that is 364 and so on down to the  $k^{\text{th}}$  person.

$$P(\text{all birthdays different}) = \frac{{}^{365}P_k}{365^k}$$

This can be evaluated on a computer for different values of  $k$ .

When  $k = 23$  one finds  $P(\text{all birthdays different}) = 0.493$ .

This implies that  $P(\text{at least one birthday the same}) = 1 - 0.493 > 0.5$ .

There is a greater than evens chance of two people having the same birthday in a room of 23 people.

### 1.4.3 Combinations

**Definition 1.16** (Combinations). The number of ways of choosing  $k$  distinct items from  $n$  when the order is not relevant is:

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

A way of choosing  $k$  distinct items from  $n$  when order does not matter is called a **combination**.

**Example 1.20.** In how many ways can 4 cards be dealt from an ordinary pack of 52 playing cards?

*solution*

Suppose one such hand is the Ace of spades, the king of clubs, the three of hearts and the Jack of diamonds. It does not matter which card you were given first, as the hand is all that matters to play.

Here ‘order does not matter’.

The number of hands is  ${}^{52}C_4 = 270725$ .

**Example 1.21.** In how many ways can 4 cards be dealt from an ordinary pack of 52 playing cards?

*solution*

Suppose one such hand is the Ace of spades, the king of clubs, the three of hearts and the Jack of diamonds. It does not matter which card you were given first, as the hand is all that matters to play.

Here ‘order does not matter’.

The number of hands is  ${}^{52}C_4 = 270725$ .

**Example 1.22** (The National Lottery). In the main National Lottery draw, six numbers are chosen from 49.

- a) What is the probability of winning the jackpot on the lottery (i.e. all 6 match)?
- b) What is the probability that three of the winning numbers come up on a lottery ticket?

*solutions*

- a) Total number of outcomes  ${}^{49}C_6 = 13983816$ .

The probability is  $\frac{1}{{}^{49}C_6}$ , which is about 1 in 14 million.

- b) The three winning numbers can be any three of the six winning numbers with  ${}^6C_3$  combinations. The other numbers on the ticket can be any three from the 43 losing numbers that week. The number of ways of choosing these is  ${}^{43}C_3$ .

Therefore the probability of three winning numbers is

$$P(\text{three winning numbers}) = \frac{{}^{43}C_3 \times {}^6C_3}{{}^{49}C_6} = 0.0177$$

This is approximately 1 in 56.