Probability Theory and Statistics

Malcolm Connolly

Semester 2, 2023

Contents

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 decribed 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 are a determinist, you 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 is 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] "Heads" "Heads" "Heads" "Tails" "Heads" "Tails" "Heads" "Tails" "Tails"
## [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:

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

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 Ω.

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

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

Definition 1.4. An ***event*** is a subset of the sample space Ω.

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 Ω.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 θ 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 Ω 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 $\varnothing$ is the set of no elements. As sets A and B are called disjoint if they have no elements in common, that is,

$A \cap B = \varnothing.$

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 $A \cap A^c = \varnothing$. 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: $\mathrm{P}(\Omega) = 1$.

(ii) *(additivity)* For any collection of disjoint events $A_1, A_2, A_3, \ldots$ the probability of the union is the sum of the probabilities. In the notation this can be written as

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots.$$

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,

$$\mathrm{P}(\Omega) = \mathrm{P}(A) + \mathrm{P}(A^c)$$

Now by axiom (i) the LHS is 1. $\qquad\square$

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,

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

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,

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

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

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?

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

$$= 1 - \frac{1}{8} \tag{1.2}$$

$$= \frac{7}{8} \tag{1.3}$$

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,

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

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:

$A_1 B A_2 A_3, A_1 B A_3 A_2, A_2 B A_1 A_3, A_2 B A_3 A_1, A_3 B A_1 A_2, A_3 B A_2 A_1$.

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, ABBAA, 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, FG 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

$$^{n}\mathrm{P}_k = \frac{n!}{(n-k)!}$$

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.

$$^{10}\mathrm{P}_4 = \frac{10!}{(10-4)!} = \frac{10 \times 9 \times \ldots 2 \times 1}{6!} = 10 \times 9 \times 8 \times 7 = 5040$$

The expression $10 \times 9 \times 8 \times 7$ can be interpreted as saying there are 10 choices for the first digit, 9 or 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

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

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 kth person.

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

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

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

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

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 (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

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

This is approximately 1 in 56.

1.5 Exercises Week 1

1.5.1 Tutorial exercises

Exercise 1.1. A letter is chosen at random from the word STATISTICS. a) What is the probability that it is a vowel? b) What is the complement of the event in a)?

Exercise 1.2. Suppose you are eating in a restaurant with two friends. You agree to pay the bill as follows. Each person tosses a coin. The person who gets a result different from the other two will pay all the bill. If all three tosses are the same, the bill will be shared equally. Find the probability that:

a) Only you will pay the bill
b) All three will share the bill

Do you think this is a *fair* way to split the bill?

Exercise 1.3. An investment can either; increase in value (I), break even (B) or make a loss (L). Suppose each outcome is equally likely. If two separate investments are made,

a) List the sample space by drawing a tree diagram.

b) Find the probability that:

(i) both investments increase in value.
(ii) both investments make a loss.
(iii) At least one of the investments increases in value.

c) Suppose both investments were in the same type of company. How might this model be unrealistic, and how could you improve it?
d) How big would the sample space be if three separate investments were made?

Exercise 1.4. A set of cards consists of the standard suits ♣, ♠, ◇, ♡, with 13 cards in each suit. a) Suppose one card is drawn at random. Find the probability that it is a: (i) Ace of Hearts, $A\heartsuit$ (ii) The King of Spades $K\spadesuit$. (iii) Any picture card.

b) Suppose two cards are drawn at random, but with the first being replaced and the deck shuffled before the second is drawn (this is called sampling with replacement). Find the probability that:

(i) Both cards are the King of Hearts, $K\heartsuit$.
(ii) Both cards are Aces.

Exercise 1.5. Fifty male and fifty female students were asked whether they agreed with the statement "Statistics are often misleading". Seventy students, thirty of whom were male, agreed. a) Summarise this information in a two-way table. b) If a student is selected at random, find the probability that they: (i) Agree (ii) Are female (iii) Are male (iv) Are male and agree (v) Are female or agree

Exercise 1.6. Interviews with 120 working people revealed that 76 were stressed, 20 were managers and 14 were both managers and stressed. a) Summarise this information in a two-way table. b) Assuming an individual is drawn at random, find the probability thatthey are (i) Stressed (ii) A shopfloor worker (iii) A manager who is stressed (iv) A shopfloor worker or is not stressed.

Exercise 1.7. Evaluate a) $^5\text{P}_3$, b) $^7\text{P}_4$, c) $^6\text{P}_4$.

Exercise 1.8. For what value of n is the following equality true?

$$^{n+1}\text{P}_3 =^n \text{P}_4$$

Exercise 1.9. Four different Mathematics books, 5 different statistics books and 3 different computing books are to be arranged on a shelf. In how many ways can the books be arranged if, a) The books in each subject must stand together b) Only the statistics books must stand together

Exercise 1.10. Four different Mathematics books, 5 different statistics books and 3 different computing books are to be arranged on a shelf. In how many ways can the books be arranged if, a) The books in each subject must stand together b) Only the statistics books must stand together

Exercise 1.11. Evaluate a) $^7\text{C}_6$, b) $^5\text{C}_3$, c) $^9\text{C}_5$, $^9\text{C}_4$.

Exercise 1.12. How many different committees can be formed from 8 men and 6 women if the committee consists of: a) 1 man and 4 women b) 5 men and 3 women c) 4 men and 4 women d) An equal number of men and women.

Exercise 1.13. A council consists of 10 members, 6 from Party X and 4 from Party Y. a) In how many ways can a committee of 4 be formed? b) In how many ways can a committee of 4 be formed so that: i) Party X has the majority ii) Party Y has the majority iii) Neither party has the majority

Exercise 1.14. Ten equally qualified assistant managersare lined up for promotion. Seven are men and three are women. If the company promotes four of the ten at random, what is the probability that exactly two of the four chosen are women?

Exercise 1.15. Suppose a library bookshelf contains an equal number, n each say, of Mathematics books and Physics books. If the bookshelf is emptied and the books placed back randomly, what is the probability that the books for each subject are separated?

Exercise 1.16. Here are some miscellaneous questions on permutations and combinations: a) From a group of 20 employees, 4 are chosen for promotion. In how many ways can they be chosen? b) From a group of 20 employees, 4 are shosen for promotion, but each to a different role. In how many ways can they be chosen? c) A product code consists of 4 letters followed by 3 digits. How many codes are possible if repetitions are not allowed? d) A 7-card hand

is dealt from a normal pack of 52 cards. How many hands will contain 4 clubs
and 3 hearts? e) How many ways can merit awards be allocated to a group of 15
students if there is one first prize, one second prize and 4 identical third prizes?
f) Four students are to be chosen from a group of 10. If exactly **one** of the first
three students must be chosen, how many ways are there of choosing the four
students?

Exercise 1.17. In the game of poker, five cards from a standard deck of 52
cards are dealt in a hand. Find the probability that a hand contains, a) A royal
flush (ace, king, queen, jack and 10 of the same suit) b) Four of a kind (e.g. all
four 5s) c) Two pairs d) A full house (i.e. three of one kind and two of another)
e) One pair

Exercise 1.18. If $P(A) = 0.6$ and $P(B) = 0.5$, can A and B be mutually
exclusive?

Exercise 1.19. The medical records of 100 male diabetic patients reported to a
clinic their family history of diabetes (Yes or No), together with their symptoms
as either mild or severe. This provided the following classification.

Age	Mild and Yes	Mild and No
under 40	15	10
40 or over	15	20

Age	Severe and Yes	Severe and No
under 40	8	2
40 or over	20	10

Suppose a patient is chosen at random from this clinic and the events A, B and
C are defines as follows:

A : He has a severe disease

B : He is under 40

C : His parents are diabetic

a) Find the probabilities $P(A)$, $P(B)$, $P(A \cap B)$, $P(B \cap C)$, $P(A \cap B \cap C)$.

b) Describe the following events in words and calculate them: $A^c \cap B^c$, $A^c \cup C^c$,
$A^c \cap B^c \cap C^c$.

1.5.2   Exercises for feedback

1. I cannot remember a phone number. It contains the following digits and is something like 132 747 6965.

   a) What is the probability that the first number is even?

   b) How many ways can the numbers above be rearranged?

   c) In how many ways can the number be rearranged to start and end with an odd number?

Suppose I am certain of the numbers in each of the blocks 132,747 and 6965, but not am not sure of the order within each block.

   c) How many ways can the numbers be rearranged such that the numbers within each block are the same?

   d) What is the probability that I wrote down the correct number originally?

2. In a lottery, 6 numbers are drawn from the numbers 1 to 49. Calculate the following probabilities.

   a) The numbers 1, 2, 3, 4, 5, 6 are all drawn.

   b) The numbers 4, 23, 24, 35, 40, 45 are all drawn.

   c) 44 is one of the numbers drawn.

3. Three dice are rolled. The sum of the numbers on the dice is the score.

   a) Describe the sample space.

   b) How many ways could the score equal 5?

   c) What is the most likely score?

4. Suppose we have a finite set S of size n. (Hint: this question is general, but you could check your answers with concrete example S = { a,b,c,d })

   a) How many subsets are there of S?

   b) How many subsets of S are there of size 1?

   c) How many subsets of S are there of size k, where $1 \le k \le n$

   d) Using a) and c), describe in words why the following equality holds.

$$2^n = \sum_{k=0}^{n} {}^nC_k$$

5. Five office workers write their names on a piece of paper, fold the paper and put them in a hat. The names are mixed up and each person then selects a piece of paper from the hat. After everyone has selected a piece of paper from the hat, the staff look at the names drawn. What is the probability that no member of staff selected their own name?

Chapter 2

Conditional Probability

In this chapter we will learn about conditional probability. This is the probability of an event, in the context of another event having happened or potentially happening.

2.1 Independence

Independence is a very important concept in Statistics, but one that is sometimes misused when it is assumed without justification. The basic idea is as follows:

Definition 2.1 (Independence)**.** Two events A and B are ***independent*** exactly when

$$P(A \cap B) = P(A) \times P(B).$$

In words this means the probability that both A and B happen is the product of the individual probabilities of A and B respectively.

Example 2.1. Some events that can be modelled as ***independent*** include: - Outcomes on successive tosses of a coin or die. What happened on the previous throw does not affect what happens on subsequent throws.

- The sex of babies. The sex of each baby is determined at random, notwithstanding the sexes of previous babies.

Example 2.2. Suppose a power plant has two safety systems, a primary system which works with probability 0.999, and a backup system which works with probability 0.89 Assuming that the two systems operate independently, what is the reliability or safety of the power plant.

solution

We can work out P(plant safe) using the complement:

$$\text{P(plant safe)} = 1 - \text{P(plant fails)}.$$

Let F be the event that the plant fails, F_1 the event that the first system fails, and F_2 the backup fails.

Then $F = F_1 \cap F_2$.

$$
\begin{align}
\text{P}(F) = \text{P}(F_1 \cap F_2) \tag{2.1} \\
= \text{P}(F_1) \times \text{P}(F_2) \tag{2.2} \\
= (1 - 0.999) \times (1 - 0.89) \tag{2.3} \\
= 0.00011 \tag{2.4}
\end{align}
$$

Then $1 - 0.00011 = 0.99989$.

Calculations such as these have often been used to arrive at unrealistic figures for the safety of complex operating processes, e.g. nuclear power plants. For example, it's easy to check that with three backup systems each with a reliability of 0.99, the probability of failure assuming independence is 1×10^{-6} - a reassuringly small figure! However we can only make calculations *if* we can justify the assumption of independence. For example it's not unusual to find that backup systems that are not used very often can be more unreliable than supposed when actually called upon.

You might have to give a reason why a particular context is not a good example in which to assume independence. For example **exercise** 1.3 part (c) asks why two investments may not be independent. There are many reasonable answers. Similar companies are dependent - if the companies are both bakeries, they may both be affected by the price of wheat. The companies may be competitors, in which case one company doing better may cause the other to do worse.

Example 2.3. Suppose you toss ten coins and coin how many are Heads. You could throw them all simultaneously. Or you could throw them one at a time, in some order. Does it matter?

solution

No, as these are independent coins. Let

$$A_i = \{\text{The } i^{\text{th}} \text{ coin is Heads}\}$$

The probability that they are simultaneously all Heads is the product of all the probabilities of each individual coin being Heads. Notice that the order does not matter as

$$\text{P}(A_i) \times \text{P}(A_j) = \text{P}(A_j) \times \text{P}(A_i).$$

Assuming independence allows us to consider simultaneous events separately one after another, complicated examples can be analysed easily using tree diagrams. Each path of a tree diagram from the root to the leaf is a distinct outcome of the sample space.

Example 2.4. Vehicles approaching a crossroads must go in one of three directions - left, right or straight on. Observations by traffic engineers showed that of vehicles approaching from the north, 45% turn left, 20% turn right and 35% go straight on. Assuming that the driver of each vehicle chooses direction independently, what is the probability that of the next three vehicles approaching from the north:

   a) all go straight on

   b) all go in the same direction

   c) two turn left and one turns right

   d) all go in different directions

   e) exactly two turn left.

solution

   a) 0.35^3

   b) $0.45^3 + 0.2^3 + 0.35^3$

   c) LLR can be rearranged in 3 ways: LLR, LRL, RLL. $3 \times 0.45^2 \times 0.2$.

   d) SRL can be rearranged in 3! ways. $3! \times 0.35 \times 0.45 \times 0.2$.

   e) LLR or LLS. Each can be rearranged in 3 ways, then these are mutually exclusive outcomes so we can add the probabilities.

$$3 \times 0.45^2 \times 0.2 + 3 \times 0.45^2 \times 0.35$$

   .

2.2 Conditional Probability

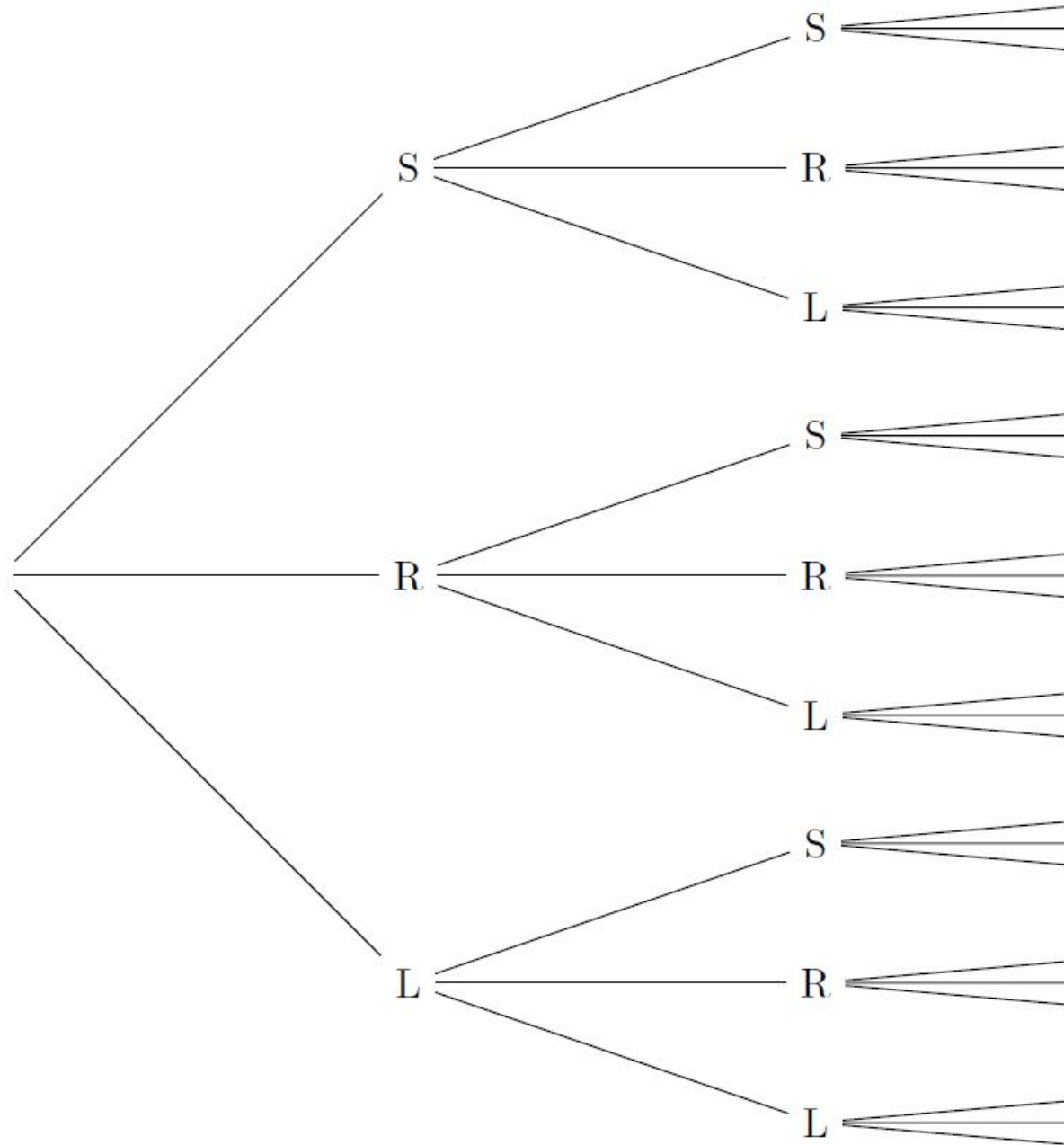We will consider the following examples to motivate the definition of conditional probability.

Figure 2.1: A tree diagram representing the choices for the three vehicles

Example 2.5. The number of insurance claims in the previous 12 months is cross tabulated with whether the driver involved was a young driver.

	Under 25	25 and over	Total
No claim	225	725	950
Claim	25	25	50
	250	750	1000

The insurance company is interested in the claim rate. Overall the claim rate is,

$$\text{P(Claim)} = \frac{50}{1000} = 0.05$$

An estimate for the probability of a driver claiming on the insurance is then 1 in 20.

However this figure hides a substantial difference in the claim rates for young and older drivers.

If we consider the 250 young drivers separately we have,

$$\text{P(Claim|Under 25)} = \frac{25}{250} = 0.1.$$

Whereas for the 750 older drivers we have,

$$\text{P(Claim|25 and over)} = \frac{25}{750} = 0.03.$$

The notation | is read 'given that' and is a conditional statement. The conditional probabilities show that the claim rate is much higher for the younger drivers. One can compute the ratio of these probabilities to see how many times higher it is, $0.1/0.03 \approx 3.3$, so this is just over three times higher. This relative risk scoring is common in medical statistics.

Example 2.6. Consider the following data from a study on male lung cancer patients carried out in 1950 in the UK. This was one of the earliest applications of epidemiology - the use of statistics to study disease patterns in populations.

	Non-smoker	Smoker	Total
Lung cancer	2	647	649
No lung cancer	27	620	647
	29	1267	1296

Calculate the relative risk of having lung cancer for a smoker compared to a non-smoker.

solution

$$\text{P(Lung cancer|Smoker)} = \frac{647}{1267}$$

$$\text{P(Lung cancer|Non-smoker)} = \frac{2}{29}$$

There is ≈ 7.4 times higher relative risk of lung cancer in smokers.

These examples motivate the definition of conditional probability.

Definition 2.2 (conditional probability)**.** The ***conditional probability*** $\text{P}(A|B)$ of an event A given another event of non-zero probability B is given by,

$$\text{P}(A|B) = \frac{\text{P}(A \cap B)}{\text{P}(B)}.$$

One should verify that the fraction on the left is precisely how the conditional probability was calculated in the previous two examples.

Theorem 2.1. *The conditional probability $P(A|B)$ satisfies Kolmogorov's definition of probability.*

Proof. Not lectured or examined, but here for completeness.

Firstly need to check $P(A|B) \in [0, 1]$. We have $P(A|B) \geq 0$ because $P(A \cap B) \geq 0$ and $P(B) > 0$.

Because the intersection of B with another set is contained in B, we have $A \cap B \subseteq B$, and so
$$P(A \cap B) \leq P(B).$$

And dividing through by $P(B)$ gives $P(A|B) \leq 1$.

Secondly,
$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Lastly, any given any two disjoint A_1, A_2 such that $A_1 \cap A_2 = \varnothing$.

We have that

$$P(A_1 \cup A_2|B) = \frac{P((A_1 \cup A_2) \cap B)}{P(B)} \tag{2.5}$$

$$= \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \tag{2.6}$$

$$= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} \quad = P(A_1|B) + P(A_2|B) \tag{2.7}$$

$\square$

Example 2.7. Note that $P(A|B) \neq P(B|A)$. Revisiting the driver's example gives,

	Under 25	25 and over	Total
No claim	225	725	950
Claim	25	25	50
	250	750	1000

$$\mathrm{P(Claim|Under\ 25)} = 0.1.$$

However,

$$\mathrm{P(Under\ 25|Claim)} = \frac{25}{50} = 0.5$$

Theorem 2.2. *Two events A and B are* independent *if and only if*

$$P(A|B) = P(A) \quad or \quad P(B|A) = P(B)$$

In other words, conditioning on either event does not affect the probability of the other event occurring.

Proof. Using the definition of conditional probability,

$$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A)$$

If

$$\mathrm{P}(A|B) = \mathrm{P}(A) \quad or \quad \mathrm{P}(B|A) = \mathrm{P}(B),$$

substituting this in the former yields

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B),$$

which is the definition of independence. Conversely if two events are independent, we have

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A),$$

and likewise for $\mathrm{P}(B|A)$. $\square$

When constructing tree diagrams the probabilities involved are usually conditional probabilities as there is a natural progression through the tree from left to right conditioning on what happened previously. In the diagram below, the events A and B may not be independent.

Example 2.8. Jon always goes to campus by bike or takes a tram. If one day he goes to campus by bike, the probability that he goes to campus by tram the next day is 0.4. If one day he goes to campus by tram, the probability that he goes to campus by bike the next day is 0.7. Given that Jon goes to campus on Monday by tram, find the probability that he takes a tram to campus on Wednesday.

solution

This may be solved by considering a tree diagram with levels for Tuesday and Wednesday. The probabilities in the question are P(tram | bike) = 0.4 and P(bike | tram) = 0.7. Monday's journey is done. Possible sequences are 'tram then tram', or 'bike then tram'. These are mutually exclusive outcomes. The calculation is then

$$0.3^2 + 0.7 \times 0.4 = 0.37$$

.

Surveys with questions of a sensitive or delicate nature often result in respondents missing that question or lying about their answers. Conditional probability can be used to mask the awkward question and find the proportion who would answer a certain way.

Example 2.9. A company want to find the proportion of employees who have ever called in sick to work, when in fact they were not sick. The boss asks each employee to toss a coin and hide the result.

If the result is **_heads_**, the employee should answer the question 'is your age an odd number?'.

If the result is **_tails_**, they should answer 'Have you ever taken a day off when you should not have?'.

Because the boss does not know which question people are answering, the employees can answer truthfully.

Suppose that 40% of employees mark 'yes' as their answer. Let,

$$p = \text{P(taken a day off | tails)}$$

Assume that ages are randomly distributed so that the chance of an even or odd number of years old is 0.5. How can we find p?
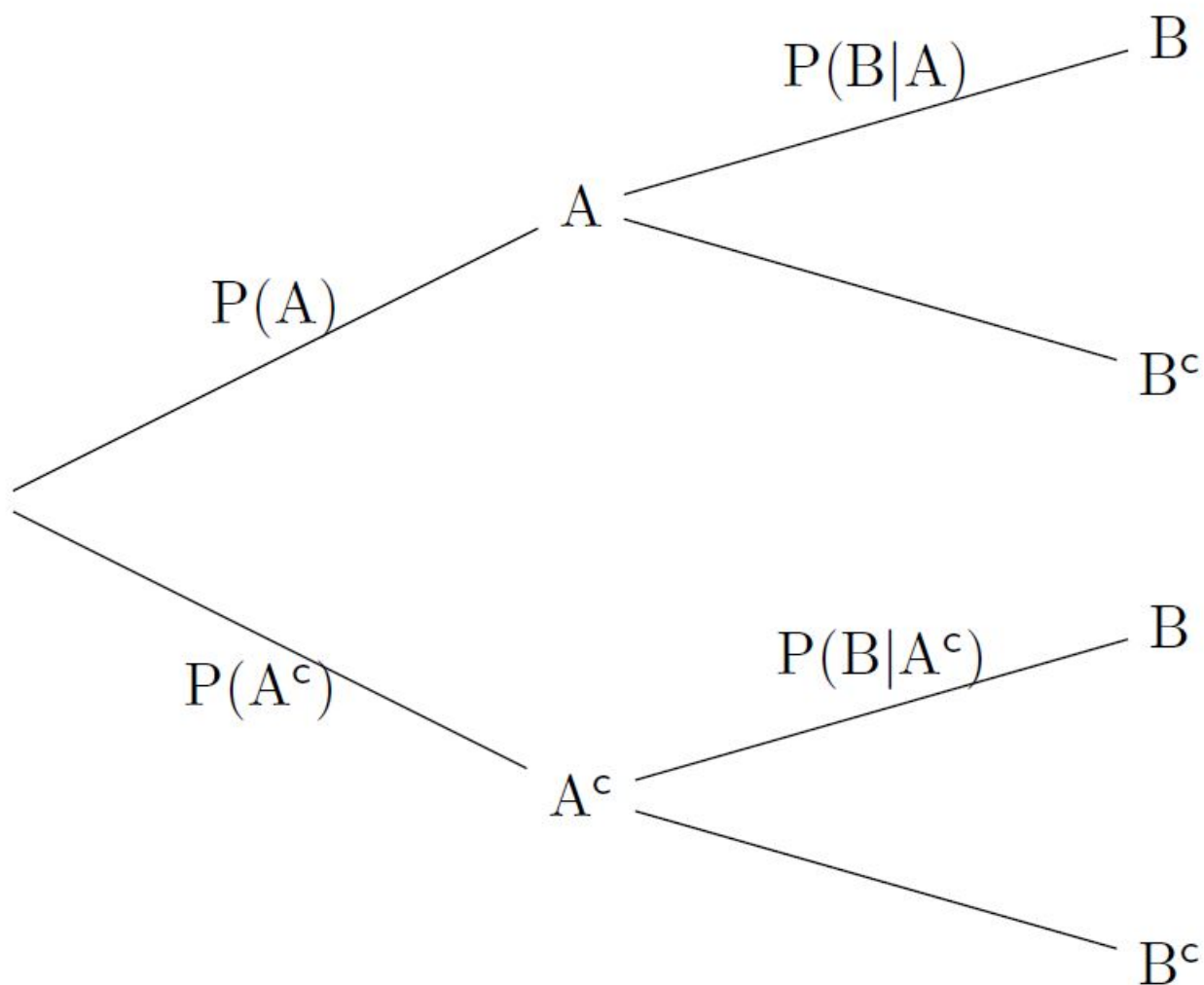
Figure 2.2: The second level of branches represent the conditional probabilities of B given A or its complement, which may be different numbers
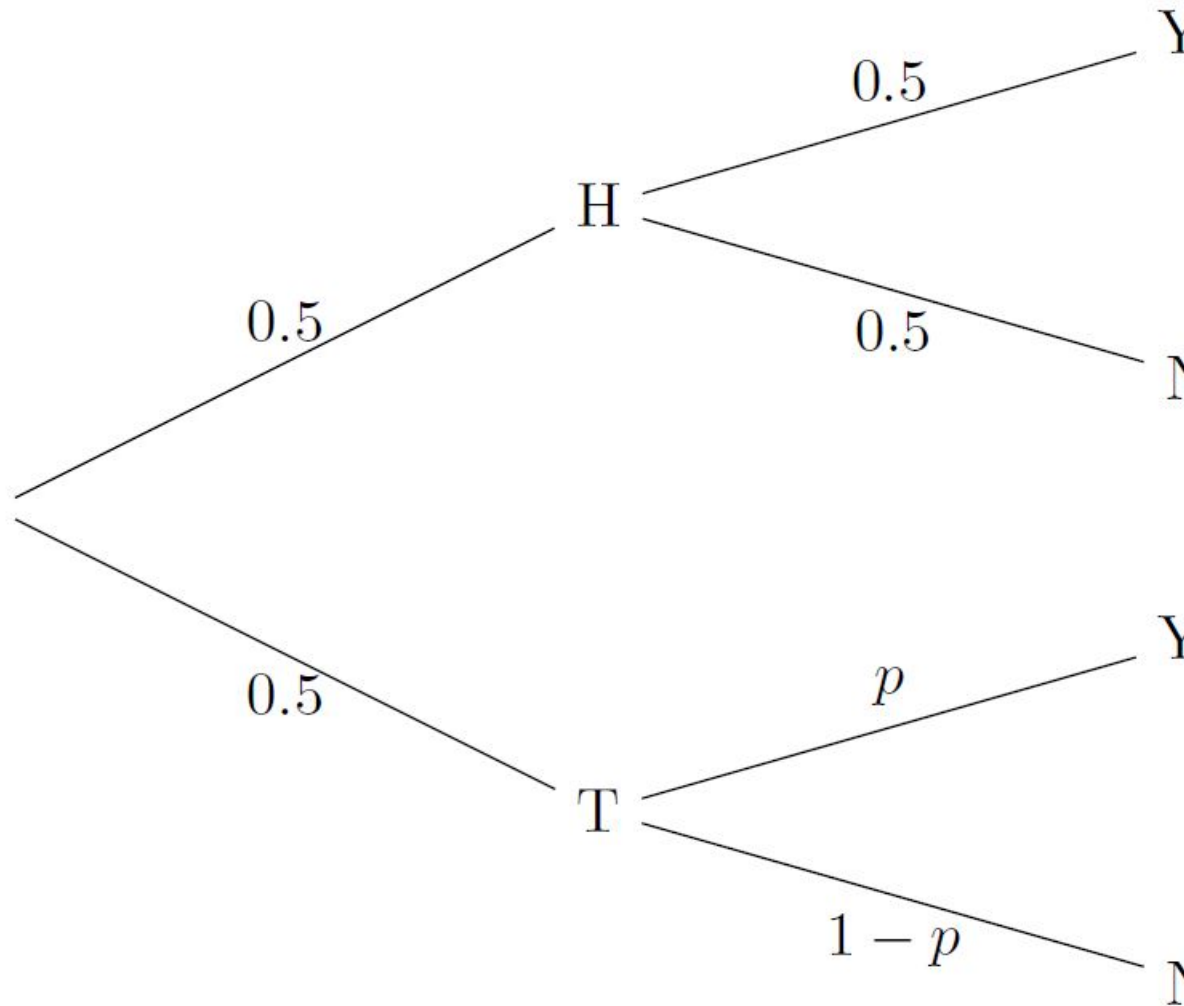
Figure 2.3: The outcomes of the survey.

solution

One can draw a tree diagram.

The overall probability of answering 'yes' is $0.25 + 0.5p$, and in the survey 40% answered 'yes'. We then have

$$0.25 + 0.5p = 0.4,$$

and hence $p = 0.3$. This means we can estimate that 30% of employees have taken a day of when they were not supposed to.

2.3 Bayes Theorem

Example 2.10. There are two coins in a bag. One coin is fair, while the other has heads on both sides (a double-header).

A coin is selected from the bag at random, and the selected coin is flipped three times. Unfortunately the coin which was selected is unknown to us.

On each of three flips the coin comes up heads.

Without doing any calculations, how likely do you think it is to be the unfair coin?

solution

Let $A = \{$The double-header is selected$\}$ and $B = \{$The coin lands heads up three times in a row$\}$

One can use the tree diagram to find $8/9$.

We can generalise this picture and come up with a formula for the conditional probability called Bayes' formula.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

Previously, $A_1 = A$ and $A_2 = A^c$ are disjoint and their union gives the entire sample space. This situation is called a *partition*.

This can be extended to a partition of n events $A_1, A_2, \ldots, A_n$.

Definition 2.3. A collection of events $A_1, A_2, \ldots, A_n$ is a ***partition*** if their union is the entire sample space, that is *exhaustive*, and they are mutually exclusive. That is

   i) $\Omega = A_1 \cup A_2 \cup \cdots \cup A_n$.

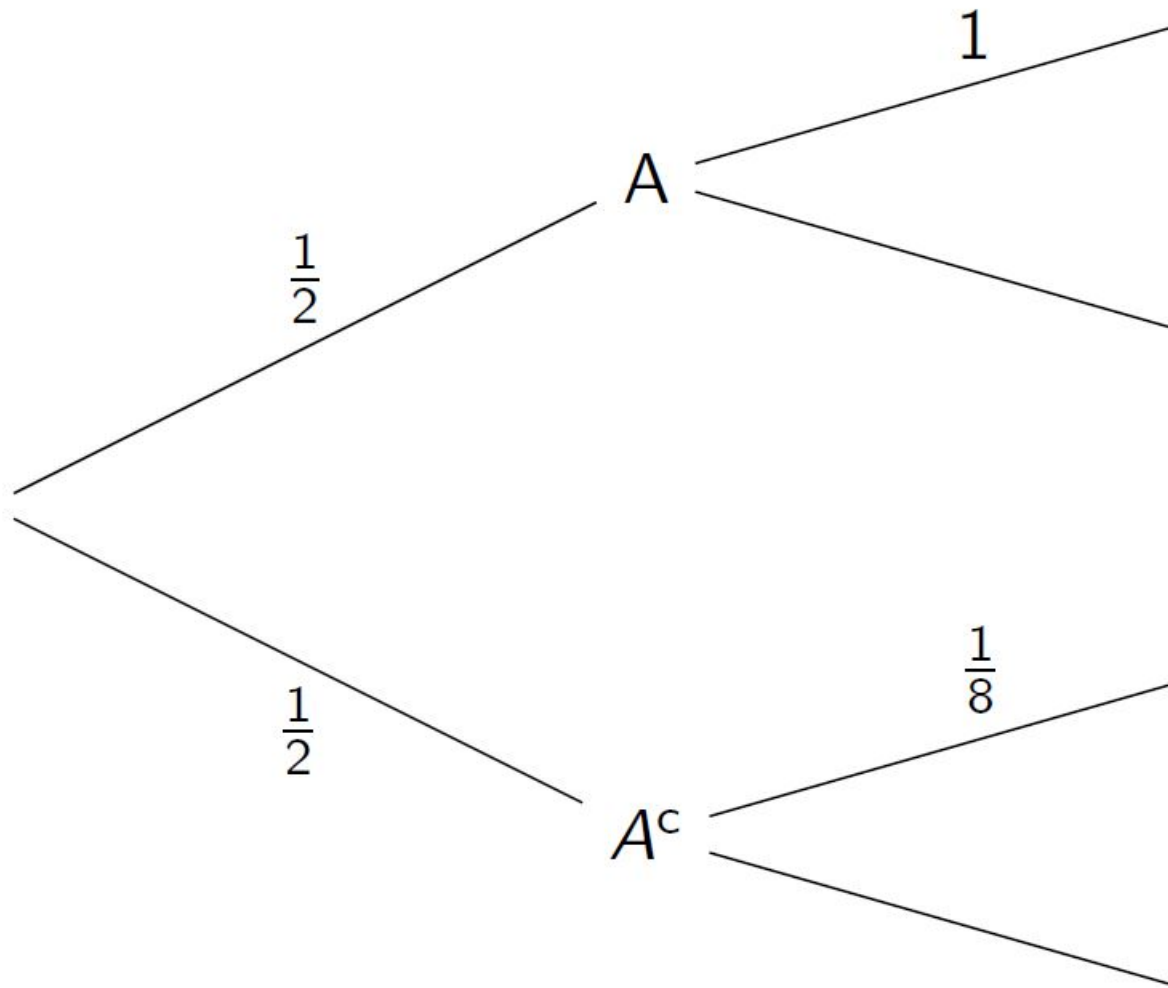   ii) $A_1 \cap A_2 \cap \cdots \cap A_n = \varnothing$

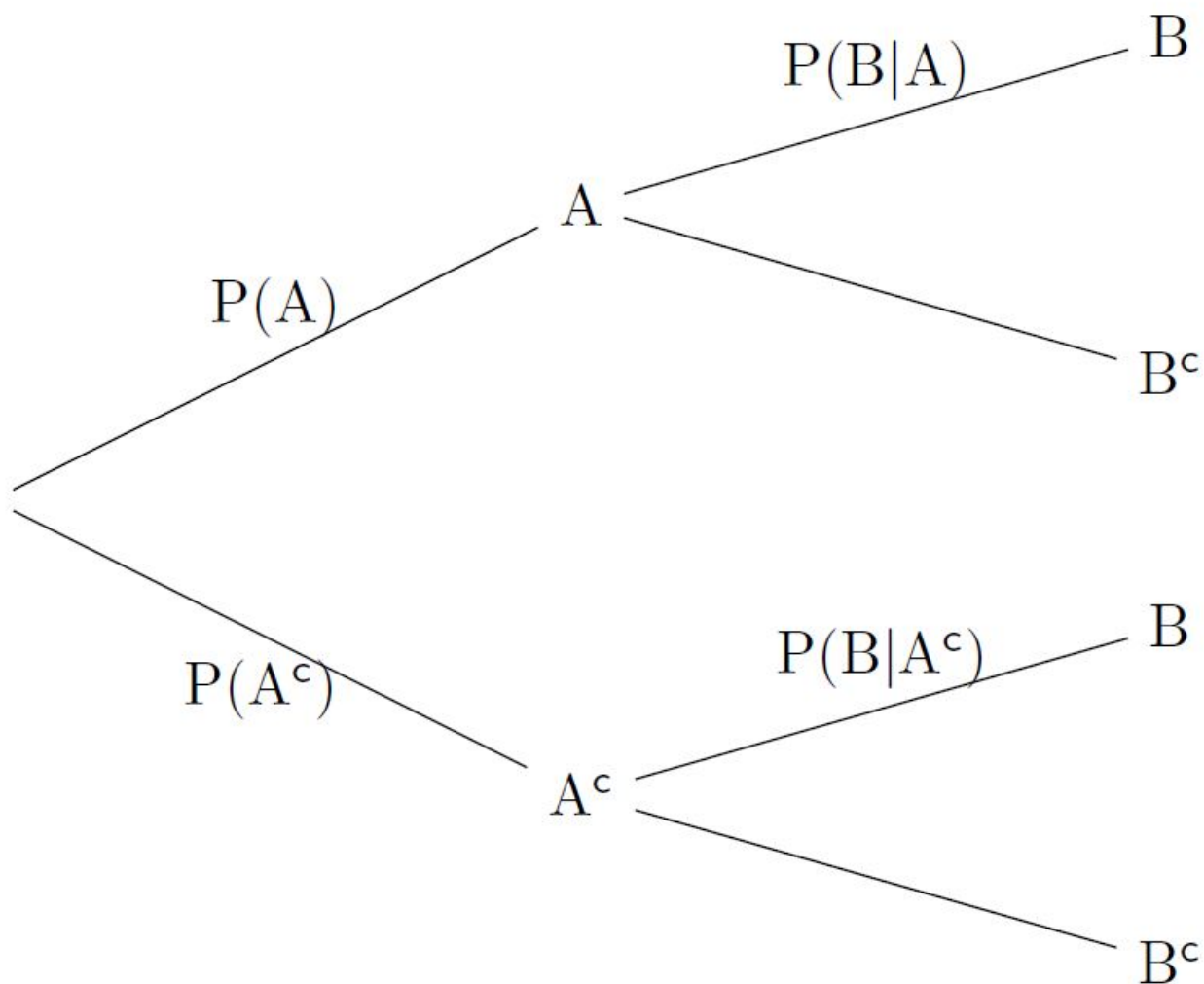Figure 2.4: A tree diagram for the double headed coin example.

Figure 2.5: Tree showing Bayes' formula

Any event and its complement form a partition.

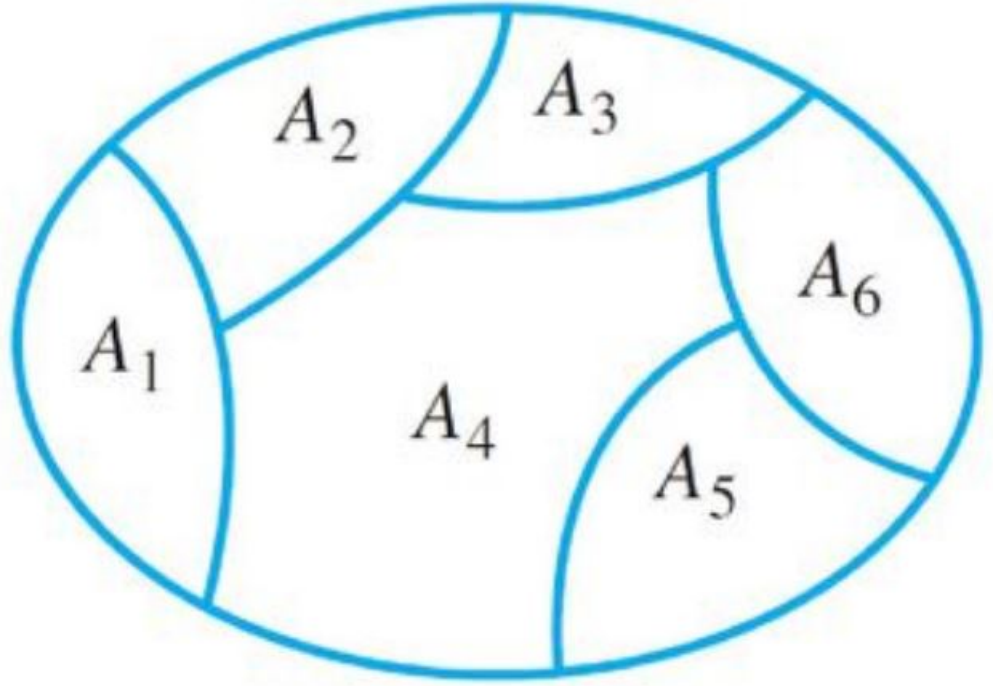Here is a picture of a partition:

Figure 2.6: An example partition with five sets.

We can now extend the concept of conditional probability to a general situation in which we condition on the event of at least one event of a partition.

Theorem 2.3 (Law of Total Probability)**.** *Suppose we have a partition* $A_1, A_2, \ldots, A_n$ *of the sample space* Ω. *Then for any event* $B \subseteq \Omega$, *we have*

$$P(B) = P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)$$

An intuitive proof is to imagine a tree diagram with n branches for each of the A_i in the first layer, then B and B^{c} in the next layer. As you multiply along all the branches the ways that B can occur you end up with the sum in the RHS.

Theorem 2.4 (Bayes' Theorem)**.** *Suppose we have a partition* $A_1, A_2, \ldots, A_n$ *of the sample space* Ω. *Then the conditional probability of any one event of the partition* A_k *for some k, given any other event B can be written as,*

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Proof. Note that $P(A_k \cap B) = P(B|A_k)P(A_k)$,

and that the denominator is $P(B) using the law of total probability. □

2.4 Tutorial Exercises

Exercise 2.1. I toss a fair coin and roll a die. a) Are these events independent?

b) What is the probability I obtain a head and a 6?

Exercise 2.2. A torch uses two batteries in series. Each battery works with probability 0.95, independently of the other. Work out the probability that:

a) The torch will work.

b) Both batteries fail

c) Only one of the batteries will work.

Exercise 2.3. Whether a student gets up on time depends on whether or not he has remembered to set his alarm the night before. Some 90% of the time he remembers, the other 10% he forgets. When the clock is set, he will get up on time 95% of occasions. If it is not set, the chance he will oversleep is 35%. Use a tree diagram to find the probability that he will oversleep.

Exercise 2.4. The following data shows the distribution of male and female students on various degree courses at a university.

	Accountancy	Economics	Finance
Male	330	360	90
Female	120	390	60

Suppose a student is selected at random. Find the probability that they are,

a) female

b) studying Economics

c) male and studying Economics

d) male given that they are studying Economics

e) female given that they are studying Economics

f) studying Economics given that they are female

Are the events 'student is male' and 'studying Economics' independent?

Exercise 2.5. The following table shows the lung cancer data for females in the same 1950 study given in example 2.6.

	Non-smoker	Smoker	Total
Lung cancer	19	41	60
No lung cancer	32	28	60
	51	69	120

a) Calculate the relative risk for female smokers compared to non-smokers.

b) Can you suggest any reason for the difference in the figures between males and females?

Exercise 2.6. Two electrical components X and Y have probabilities of working $\frac{3}{4}$ and $\frac{7}{8}$, respectively. They also function independently of each other. Two devices D_1 and D_2 are constructed. In D_1, X and Y are in series, and in D_2 they are wired in parallel.

a) (i) Find the probability that D_1 works.

(ii) Find the probability that D_2 works.

b) Suppose that D_1 works, find the probability that;

(i) X is working.
(ii) Only X is working.
(iii) both X and Y are working.

c) Suppose that D_2 works, find the probability that;

(i) X is working.
(ii) Only X is working.
(iii) both X and Y are working.

Exercise 2.7. An urn contains two green balls and three red bals. Supose two balls will be drawn at random one after another and without replacement. Draw a tree diagram, and find the probability that:

a) a green ball appears on the first draw.

b) a green ball appears in the second draw.

Exercise 2.8. The following table shows the *fear factor* for children attending the dentist, cross tabulated with the School age of the child.

	Infant	Primary	Secondary
Afraid	0.12	0.08	0.05
Not afraid	0.28	0.25	0.22

For a child selected at random define the events; $A = \{\text{The child is afraid}\}$,

with N being not afraid, and I, P and S being the School age in the obvious fashion.

Calculate the following probabilities,

a) $P(A)$, $P(N)$, $P(A \cup I)$.

b) $P(A|I)$ and $P(I|A)$.

c) $P(A|S)$ and $P(N|S)$ - what do you notice about these two probabilities?

Are A and I independent?

Exercise 2.9. A survey by an electrical retailer determines that 40% of customers who seek advice from sales staff by an appliance and only 20% who do not seek advice buy an appliance. If 30% of customers seek advice, what is the probability that a customer entering the warehouse buys an appliance?

Exercise 2.10. Four cards are drawn at random without replacement from a deck of 52 cards. What is the probability that the sequence is:

a) ♡ ♡ ♠ ♣

b) ♡ ♡ ♠ ♠

Exercise 2.11. A student comes back from a night at the pub with a bunch of keys, only one of which works. They try one key at random in the lock and discard it if it doesn't fit.

a) Suppose the bunch contains 2 keys. Find the probability they open the door on

(i) the first attempt

(ii) the second attempt

b) Repeat for a bunch of three keys being successul at the first, second and third attempts.

c) Suppose now that the bunch contains n keys. Find the probability that the door is opened on the rth attempt (where $1 \le r \le n$).

Exercise 2.12. To ascertain the proportion of people who have had a sexually transmitted infection, the following survey pocedure was used on 1000 individuals.

They were asked to think of the day of the week their most recent birthday fell on.

If their last birthday was on a Monday, Tuesday or Wednesday they were to answer the question 'Have you every had a sexually transmitted infection?'.

If their last birthday was on any other day of the week, they were to answer the question 'Is your age an even number?'.

In the survey 290 people answered 'yes'. Assuming that ages and birthdays are uniformly distributed, can you estimate the proportion of people who have had a sexually transmitted infection?