

Chapter 4

Probability

What we'll cover...

- **Definitions**: Experiments, outcomes, sample space...
- **Interpretations** of probability: *Classical, Frequentist, Bayesian*
- **Laws** of probability
- Probability **trees**
- The **binomial** distribution
- The **Poisson** distribution

4. Probability

Probability is the language we use to model uncertainty.

- Few things in life are certain
- Usually an element of uncertainty around outcomes of our choices
- This can make all the difference between a good investment and a poor one!
- Hence the importance of probability!

4.1 Definitions

We often use the letter P to represent a probability. For example, $P(\text{Rain})$ would be the probability that it rains.

An **Experiment** is an activity where we do not know for certain what will happen, but we will *observe* what happens. For example:

- We will ask someone whether or not they have used our product.
- We will observe the temperature at midday tomorrow.
- We will toss a coin and observe whether it shows “heads” or “tails”.

4.1 Definitions

An **outcome**, or **elementary event**, is one of the possible things that can happen.

For example, suppose that we are interested in the (UK) shoe size of the next customer to come into a shoe shop. Possible outcomes:

- “eight”
- “twelve”
- “nine and a half”

In any experiment, **one, and only one**, outcome occurs.

4.1 Definitions

The **sample space** is the set of all possible outcomes. For example, it could be the set of all shoe sizes.

An **event** is a set of outcomes.

For example “the shoe size of the next customer is less than 9” is an event.

It is made up of all of the outcomes where the shoe size is less than 9.

4.1 Definitions

Probabilities are usually expressed in terms of fractions or decimal numbers or percentages.

Therefore we could express the probability of it raining today as

$$P(\text{Rain}) = \frac{1}{20} = 0.05 = 5\%.$$

All probabilities are measured on a scale ranging from **0** → **1**.

- An event with probability zero is an **impossible** event
- An event with probability one is a **certain** event.
- The collection of all possible outcomes, that is the **sample space**, has a probability of 1.

For example, if an event consists of only two outcomes – *success* or *failure* – then $P(\text{success or failure}) = 1$.

4.1 Definitions

Two events are said to be **mutually exclusive** if both cannot occur simultaneously: In the example above, the outcomes *success* and *failure* are mutually exclusive.

Two events are said to be **independent** if the occurrence of one does not affect the probability of the second occurring.

For example, if you toss a coin and look out of the window, it would be reasonable to suppose that the events “get heads” and “it is raining” would be independent.

4.2 How do we measure Probability?

There are three main ways in which we can measure probability:

- The **classical** interpretation
- The **frequentist** interpretation
- The **subjective** or **Bayesian** interpretation

All three obey the basic rules described above.

Different people argue in favour of the different views of probability and some will argue that each kind has its uses depending on the circumstances.

4.2.1 The classical interpretation

If all possible outcomes are “**equally likely**” then we can adopt the **classical** approach to measuring probability.

For example, if we tossed a fair coin, there are only two possible outcomes – a head or a tail – both of which are equally likely, and hence

$$P(\textit{Head}) = \frac{1}{2} \quad \text{and} \quad P(\textit{Tail}) = \frac{1}{2}.$$

4.2.1 The classical interpretation

The underlying idea behind this view of probability is **symmetry**.

In this example, there is no reason to think that the outcome *Head* and the outcome *Tail* have different probabilities and so they should have the same probability.

Since there are two outcomes and one of them must occur, both outcomes must have probability $1/2$.

Generally,

$$P(\text{Event}) = \frac{\text{Total number of outcomes in which event occurs}}{\text{Total number of possible outcomes}}.$$

4.2.2 The frequentist interpretation

When the outcomes of an experiment are **not** equally likely, we can **conduct experiments** to give us some idea of how likely the different outcomes are.

For example, suppose we were interested in measuring the probability of producing a defective item in a manufacturing process.

- Monitor the process over a long period
- Calculate the proportion of defective items

What constitutes a “long period of time”?

Generally,

$$P(\text{Event}) = \frac{\text{Number of times an event occurs}}{\text{Total number of times experiment done}}.$$

4.2.3 The Subjective/Bayesian interpretation

We are probably all intuitively familiar with this method of assigning probabilities:

- When we board a plane, we judge the probability of it crashing to be sufficiently small that we are happy to undertake the journey.
- Similarly, the odds given by bookmakers on a horse race reflect people's beliefs about which horse will win.

This probability does not fit within the frequentist definition as the same race cannot be run a large number of times.

Example 4.1

A fast-food chain with 700 outlets describes the geographic location of its restaurants with the following table:

		Region			
		NE	SE	SW	NW
Population	Under 10,000	35	42	21	70
	10,000–100,000	70	105	84	35
	Over 100,000	175	28	35	0

A health and safety organisation selects a restaurant at random for a hygiene inspection.

- (a) Which of the three approaches to probability would you use to find:
- (i) $P(\text{NE restaurant chosen})$,
 - (ii) $P(\text{SW and city with a population} < 100,000)$?
- (b) Now use this approach to find the probabilities given above.

Example 4.1: Solution to part (a)

A restaurant is chosen at random – all outcomes are equally likely – so **classical**.

[Not that important, really...]

Example 4.1: Solution to part (b)(i)

		Region			
		NE	SE	SW	NW
Population	Under 10,000	35	42	21	70
	10,000–100,000	70	105	84	35
	Over 100,000	175	28	35	0

$$\begin{aligned}P(\text{NE restaurant chosen}) &= \frac{35 + 70 + 175}{700} \\&= \frac{280}{700} \\&= 0.4\end{aligned}$$

Example 4.1: Solution to part (b)(ii)

		Region			
		NE	SE	SW	NW
Population	Under 10,000	35	42	21	70
	10,000–100,000	70	105	84	35
	Over 100,000	175	28	35	0

$$\begin{aligned}P(\text{SW and city with a population} < 100,000) &= \frac{21 + 84}{700} \\&= \frac{105}{700} \\&= 0.15\end{aligned}$$

4.3.1 Laws of Probability: Multiplication law

The probability of two **independent** events E_1 and E_2 both occurring can be written as

$$P(E_1 \text{ and } E_2) = P(E_1) \times P(E_2),$$

and this is known as the **multiplication law** of probability.

For example, the probability of throwing a six followed by another six on two rolls of a die is calculated as follows:

- The outcomes of the two rolls of the die are **independent**
- Let E_1 denote a six on the first roll and E_2 a six on the second roll.



$$\begin{aligned} P(\text{two sixes}) &= P(E_1 \text{ and } E_2) \\ &= P(E_1) \times P(E_2) \\ &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}. \end{aligned}$$

4.3.2 Laws of probability: Addition Law

The multiplication law is concerned with the probability of two or more independent events occurring.

The **addition law** describes the probability of any of two **or** more events occurring.

The addition law for two events E_1 and E_2 is

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ and } E_2).$$

This describes the probability of *either* event E_1 *or* event E_2 happening.

4.3.2 Laws of probability: Addition Law

A more basic version of the rule works where events are **mutually exclusive**.

If events E_1 and E_2 are mutually exclusive then

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2).$$

This simplification occurs because when two events are mutually exclusive they cannot happen together and so $P(E_1 \text{ and } E_2) = 0$.

4.5 The binomial distribution

Suppose we are interested in the **number of sixes** we get from 3 rolls of a die.

Each roll of the die is an experiment or trial which gives a “**six**” (success, or **s**) or “**not a six**” (failure, or **f**).

The probability of a success is $p = P(\text{six}) = 1/6$.

We have $n = 3$ **independent** experiments or trials (rolls of the die).

4.5 The binomial distribution

Let X be the number of sixes obtained.

We can now obtain the full **probability distribution** of X ; a probability distribution is a list of all the possible outcomes for X with along with their associated probabilities.

4.5 The binomial distribution

For example, suppose we want to work out the probability of obtaining **three** sixes: (three “**successes**” — i.e. **sss** — or $P(X = 3)$).

Since the rolls of the die can be considered independent, we get (using the **multiplication law**):

$$P(\text{sss}) = P(s) \times P(s) \times P(s) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^3$$

4.5 The binomial distribution

That one's easy!

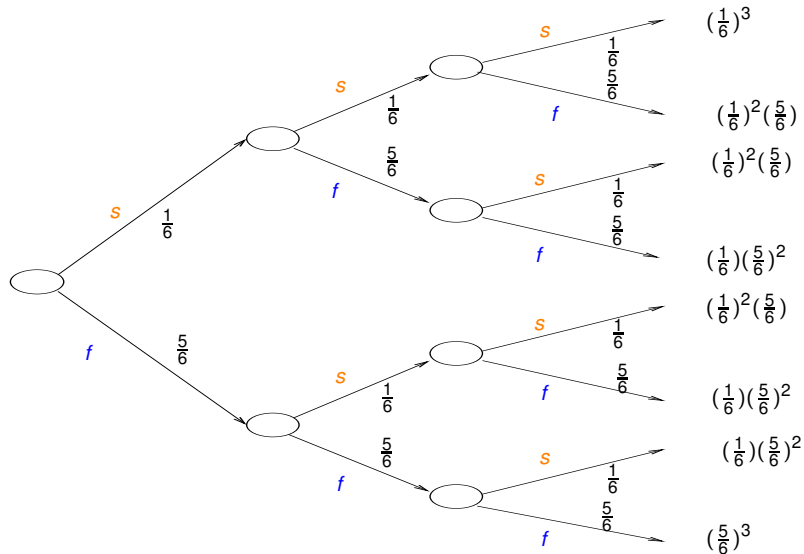
What about the probability that we get **two** sixes — i.e.
 $P(X = 2)$?

This one's a bit more tricky, because that means we need two
s's and one **f**...

...but the **f** (“not six”) could appear on the first roll, or the
second roll, or the third!

Thinking about it, there are actually **eight** possible outcomes
for the three rolls of the die:

4.5 The binomial distribution



4.5 The binomial distribution

So, for $P(X = 2)$, we *could* have:

$$P(\text{fss}) = \frac{5}{6} \times \frac{1}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^2 \times \frac{5}{6},$$

or we *could* have:

$$P(\text{sfs}) = \frac{1}{6} \times \frac{5}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^2 \times \frac{5}{6},$$

or even:

$$P(\text{ssf}) = \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} = \left(\frac{1}{6}\right)^2 \times \frac{5}{6},$$

4.5 The binomial distribution

Can you see that we therefore get:

$$P(X = 2) = 3 \times \left(\frac{1}{6}\right)^2 \times \frac{5}{6}.$$

Which takes the form:

$$\begin{aligned} P(X = 2) &= \text{Number of ways to get two sixes} \\ &\times P(2 \text{ sixes}) \times P(1 \text{ "not six"}). \end{aligned}$$

4.5 The binomial distribution

Using the same argument as above we can calculate the other probabilities:

$$P(X = 0) = \left(\frac{5}{6}\right)^3 = 0.579$$

$$P(X = 1) = 3 \times \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)^2 = 0.347$$

$$P(X = 2) = 3 \times \left(\frac{1}{6}\right)^2 \times \frac{5}{6} = 0.069$$

$$P(X = 3) = \left(\frac{1}{6}\right)^3 = 0.005...$$

4.5 The binomial distribution

... and so the full **probability distribution** for X is:

x	0	1	2	3
$P(X = x)$	0.579	0.347	0.069	0.005

This probability distribution shows that most of the time we would get either 0 or 1 sixes and, for example, 3 sixes would be quite rare.

Try your own experiment!

Minitab

4.5 The binomial distribution

Now this is a bit long-winded . . . and that was just for **three** rolls of the die!

Imagine what it would be like to calculate for **100** rolls of the die!

We would like a more concise way of working these probabilities out without having to list all the possible outcomes as we did above.

4.5 The binomial distribution

You should see from the tree diagram that we can construct a **general formula**, taking the form:

$$P(X = r) = \# \text{ ways to get } r \text{ successes out of } n \text{ trials} \\ \times P(r \text{ successes}) \times P(n - r \text{ failures})$$

4.5.1 The binomial distribution: notation

In many surveys and experiments data is collected in the form of **counts**. For example:

- The number of people in the survey who bought a CD in the past month
- The number of people who said they would vote Labour at the next election
- The number of defective items in a sample taken from a production line

4.5.1 The binomial distribution: notation

All these variables have common features:

1. Each person/item has only two possible (exclusive) responses (Yes/No, Defective/Not defective etc)
— this is referred to as a **trial** which results in a **success** or **failure**.
2. The survey/experiment takes the form of a random sample
— the responses are **independent**.
3. The probability of a success in each trial is p (in which case the probability of a failure is $1 - p$).
4. We are interested in the random variable X , the total number of successes out of n trials.

4.5.1 The binomial distribution: notation

If these conditions are met, then X has a **binomial distribution** with index n and probability p .

We write this as $X \sim \text{Bin}(n, p)$, which reads as “ X has a binomial distribution with index n and probability p ”.

Here, n and p are known as the **parameters** of the binomial distribution.

4.5.1 The binomial distribution: notation

Returning to the dice-rolling experiment earlier: we were interested in the number of sixes obtained from three rolls of a six-sided die.

Treating each roll of the die as a **trial**, with a six representing a **success** and “not a six” representing a **failure**, we can see that:

- We have $n = 3$ independent trials;
- Each trial has probability of success $p = 1/6$.

Thus if X represents the number of sixes on the three rolls, we have that

$$X \sim \text{Bin}(3, 1/6).$$

4.5.2 Probability calculations

Recall our derived formula for working out probabilities in the die example:

$$P(X = r) = \text{\# ways to get } r \text{ successes out of } n \text{ trials} \\ \times P(r \text{ successes}) \times P(n - r \text{ failures}).$$

We can write this more succinctly as

$$P(X = r) = {}^nC_r \times p^r \times (1 - p)^{n-r}.$$

4.5.2 Probability calculations

$$P(X = r) = {}^nC_r \times p^r \times (1 - p)^{n-r}.$$

The **binomial coefficient** ${}^nC_r$ works out how many ways we can choose r objects out of n , and so is commonly read as “ n choose r ”.

There is a formula for this, but it is more easily obtained using the nCr button on the calculator!

4.5.2 Probability calculations

$$P(X = r) = {}^nC_r \times p^r \times (1 - p)^{n-r}.$$

The expression for $P(X = 2)$ in the die rolling example can be calculated using this formula.

In this example,

- $n = 3$
- $r = 2$
- The probability of success is $p = \frac{1}{6}$
- The probability of failure is $1 - p = \frac{5}{6}$.

4.5.2 Probability calculations

So,

$$\begin{aligned}P(X = 2) &= {}^3C_2 \times \left(\frac{1}{6}\right)^2 \times \left(1 - \frac{1}{6}\right)^{3-2} \\&= 3 \times \left(\frac{1}{6}\right)^2 \times \frac{5}{6} \\&= 0.069,\end{aligned}$$

which matches up exactly with what we calculated before (as it should)!

You should try this formula for $r = 0, 1$ and 3 to make sure you can recover the other values in the table on ~~page 112~~ page 104.

4.5.3 Mean and variance

If we have the probability distribution for X rather than the raw observations:

- We denote the mean for X not by $\bar{x}$ but by $E[X]$ (“expectation of X ”)
- We denote the variance by $Var(X)$, and the standard deviation by $SD(X)$

If X is a random variable with Binomial $Bin(n, p)$ distribution, then its mean and variance are

$$E[X] = n \times p \quad Var(X) = n \times p \times (1 - p).$$

4.5.3 Mean and variance

For example, in the die-rolling experiment,

$$E[X] = 3 \times \frac{1}{6} = 0.5$$

and

$$\text{Var}(X) = 3 \times \frac{1}{6} \times \frac{5}{6} = 0.417 \quad \text{and so}$$

$$\text{SD}(X) = \sqrt{0.417} = 0.645.$$

Example 4.4

A salesperson has a 40% chance of making a sale on a customer visit and she arranges 6 visits in a day. Suppose that the salesperson's visits result in sales *independently*.

- (a) What are the probabilities of her making 0, 1, 2, 3, 4, 5 and 6 sales?
- (b) Find the salesperson's expected number of sales. What is the standard deviation?

Solution to Example 4.4 (a)

Let X : number of sales. Then we have that

$$X \sim \text{Bin}(6, 0.4)$$

Thus

$$P(X = 0) = {}^6C_0 \times 0.4^0 \times 0.6^6 = 0.047$$

$$P(X = 1) = {}^6C_1 \times 0.4^1 \times 0.6^5 = 0.187$$

$$P(X = 2) = {}^6C_2 \times 0.4^2 \times 0.6^4 = 0.311$$

$$\vdots \quad \vdots \quad \vdots$$

$$P(X = 6) = {}^6C_6 \times 0.4^6 \times 0.6^0 = 0.004.$$

Solution to Example 4.4 (b)

If $X \sim \text{Bin}(n, p)$,

- $E[X] = n \times p$ and

- $\text{Var}(X) = n \times p \times (1 - p)$,

giving

$$E[X] = 6 \times 0.4 = 2.4 \text{ sales.}$$

Also,

$$\text{Var}(X) = 6 \times 0.4 \times 0.6 = 1.44 \quad \text{so}$$

$$\text{SD}(X) = \sqrt{1.44} = 1.2 \text{ sales.}$$