Probability

Probability can loosely be described as the proportion of times, out of a given number of attempts, that a particular event is **expected** to happen. This does not mean that the particular event will definitely happen that number of times.

We will adopt a simple definition which should become clearer as we go through the material. First of all, we will introduce some simple terminology.

- An **experiment** is the specification of an activity which will allow the observation of something of interest, e.g., flipping a coin once or requesting six students simultaneously to log in to the network.
- A **trial** is one practical implementation of the experiment.
- An **outcome** is the result of one particular trial.
- An **event** is the outcome or set of outcomes (subset of all possible outcomes) in which one is interested.

An event is usually defined by a **statement** which, in a particular trial, will either be true or false. **Success** in a trial means that the desired outcome is achieved, i.e., the corresponding **statement** is true. E.g., getting a head (because that was what was wanted) when flipping a coin. **Failure**, of course, is the opposite of success.

Now we can give a simple interpretation of probability:

- The average proportion of successes over many trials approaches the probability (value) as the number of trials increases.

This interpretation is very important. It might not be easy to understand now but an appreciation of it should be developed.

For example, when the experiment is the tossing of a fair 2-sided coin several times, we would **expect** half of the outcomes to be a head uppermost and half of the outcomes to be a tail uppermost even though this is not the only possibility.

Hence, when a coin is tossed once, the probability that a head will show is $\frac{1}{2}$. In addition, the corresponding probability that a tail will show is also $\frac{1}{2}$. There are no other possible outcomes here and the two outcomes cannot occur simultaneously. Therefore, these 2 probabilities add up to 1. The more times we flip the coin the more likely we will get half of the results to be head uppermost.

If our experiment were to flip the coin 8 times then we would expect 4 heads. If we carry out many trials of this experiment then the average number of heads per trial will be close to 4 and is likely to get closer to 4 as we increase the number of similar trials.

As a second example, the relative frequencies calculated from the frequencies in the table for the grouped data in the table below may be interpreted as probabilities. In particular, this table gives the probability that one of the 100 people selected at random is in the 15 to 19 (not yet 20) years old group. This probability is $\frac{6}{100}$, i.e., 0.06.

Age Range	**Number in Group (Frequency)**
$0 \leq x < 5$	1
$5 \leq x < 10$	4
$10 \leq x < 15$	12
$15 \leq x < 20$	6
$20 \leq x < 25$	5
$25 \leq x < 30$	5
$30 \leq x < 35$	7
$35 \leq x < 40$	5
$40 \leq x < 45$	6
$45 \leq x < 50$	9
$50 \leq x < 55$	8
$55 \leq x < 60$	6
$60 \leq x < 65$	8
$65 \leq x < 70$	5
$70 \leq x < 75$	7
$75 \leq x < 80$	3
$80 \leq x < 85$	3
Total:	100

This relative frequency approach leads to a very simple definition of probability. Let A represent a statement which will be either True or False. If we carry out several trials then in some trials A will be true and in the other trials A will be false. Let there be N such trials and let S of these result in A being true, i.e., a success. Then, on any such single trial, the **probability that A is true (or probability of success) is calculated as**:

$$P(A) = \frac{S}{N}.$$

Obviously $0 \leq S \leq N$ and so $0 \leq P(A) \leq 1$.

Also we have the complement, or negation, of A and the corresponding probability:

$$P(\neg\ A) = \frac{N - S}{N} = 1 - \frac{S}{N}.$$

Total probability is always 1, so we have $P(A) + P(\neg\ A) = 1$.

We can use this simple definition whenever we can list all possible outcomes and count those in which our chosen statement is true.

Example 1 Two dice (normal six-sided cubic) are rolled and the sum of spots on the 2 upper faces is calculated. The possibilities range from 2 to 12 and are tabulated below. Our chosen statement could be that this sum is exactly 7 and the relevant outcomes (i.e., make the statement true) are highlighted. We can count the number of relevant outcomes and then calculate the corresponding probability.

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	**(1, 6)**
(2, 1)	(2, 2)	(2, 3)	(2, 4)	**(2, 5)**	(2, 6)
(3, 1)	(3, 2)	(3, 3)	**(3, 4)**	(3, 5)	(3, 6)
(4, 1)	(4, 2)	**(4, 3)**	(4, 4)	(4, 5)	(4, 6)
(5, 1)	**(5, 2)**	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

There are 36 possible outcomes (all equally likely) of which 6 make our statement true. Let our statement be A_1 = 'The sum is 7'. Then the probability of success is $P(A_1)$, which is

$$P(A_1) = \frac{6}{36} = \frac{1}{6}.$$

Similarly, the negation of the statement A_1 is $\neg\, A_1$ = 'The sum is not 7'. The probability of this is then

$$P(\neg\, A_1) = 1 - \frac{1}{6} = \frac{5}{6}.$$

Example 2 Following on from Example 1, we can have a statement such as A_2 = 'The sum is less than 6'. There are 4 relevant outcomes in Row 1 of the table, 3 outcomes in Row 2, 2 in Row 3, 1 in Row 4, and no more. Hence, there are 10 outcomes in which A_2 is true, and so

$$P(A_2) = \frac{10}{36} = \frac{5}{18}.$$

Example 3 When dealing with a compound statement, such as A_1 or A_2, we can define it in terms of simple statements. Let us use algebraic quantities x and y to indicate the numbers showing on the two dice. Obviously $1 \le x \le 6$ and $1 \le y \le 6$. Compound statement A_1 can be given in terms of simple statements as

$$\begin{aligned}A_1 =&[x + y = 7]\\ =&[(x = 1) \wedge (y = 6)] \vee [(x = 2) \wedge (y = 5)] \vee [(x = 3) \wedge (y = 4)]\\ &[(x = 4) \wedge (y = 3)] \vee [(x = 5) \wedge (y = 2)] \vee [(x = 6) \wedge (y = 1)].\end{aligned}$$

We have twelve simple statements, such as $(x = 1)$, or $(y = 1)$, and these are connected in pairs to form compound statements such as $[(x = 1) \wedge (y = 6)]$ and then we have six of these compound statements connected to form statement A_1.

Define the following statements. Then $A_1 = B_1 \vee B_2 \vee B_3 \vee B_4 \vee B_5 \vee B_6$. The events B_1 up to B_6 are **mutually exclusive**. That is, if any one occurs, then the others cannot occur.

$$
\begin{array}{lll}
B_1 = (x=1) \wedge (y=6) & B_2 = (x=2) \wedge (y=5) & B_3 = (x=3) \wedge (y=4) \\
B_4 = (x=4) \wedge (y=3) & B_5 = (x=5) \wedge (y=2) & B_6 = (x=6) \wedge (y=1)
\end{array}
$$

Also, in B_1 itself, the truth value of $(x = 1)$ is **independent** of the truth value of $(y = 6)$ and vice-versa. That is because the two *events* are **independent**.

If we concentrate just on the possible values of x then we can see that $P(x = 1) = \frac{1}{6}$ and similarly if we concentrate just on the values of y then we can see that $P(y = 6) = \frac{1}{6}$. If we consider the combination $B_1 = (x = 1) \wedge (y = 6)$, then the table of the previous example gives

$$P(B_1) = \frac{1}{36}.$$

We can interpret these probabilities in the following way (for a large number of rolls of two dice). Statement $(x = 1)$ will be true $\frac{1}{6}$ of the time and $(y = 6)$ will be true $\frac{1}{6}$ of those times when $(x = 1)$ is true. Hence $(x = 1)$ and $(y = 6)$ are simultaneously true $\frac{1}{36}$ of the time. This can be summarised as

$$P((x = 1) \wedge (y = 6)) = P(x = 1) \times P(y = 6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

This is the **multiplication law of probability for independent events**.

Two events are independent if the outcome in one event does not affect the outcome in the other event (and vice-versa). In this case, to determine the probability that both events occur simultaneously, we simply multiply together the probabilities that each occurs, as in the case of B_1 above.

Returning to the details of Example 3, we can now see that, in addition to $P(B_1) = \frac{1}{36}$, we have

$$P(B_2) = \frac{1}{36}, P(B_3) = \frac{1}{36}, P(B_4) = \frac{1}{36}, P(B_5) = \frac{1}{36}, P(B_6) = \frac{1}{36}.$$

Events B_1 to B_6 lead to the diagonal entries identified in bold in the table of Example 1. In calculating $P(A_1)$ from the table in Example 1, we noted that there were 6 relevant outcomes of the total 36, giving the probability of $\frac{6}{36}$. It is no coincidence that this value is equal to the sum of $P(B_1)$, $P(B_2)$, $P(B_3)$, $P(B_4)$, $P(B_5)$, and $P(B_6)$. As noted before, the 6 events B_1 to B_6 are mutually exclusive. Hence, in order to determine that (at least) one of them occurs, we simply add up the 6 separate probabilities:

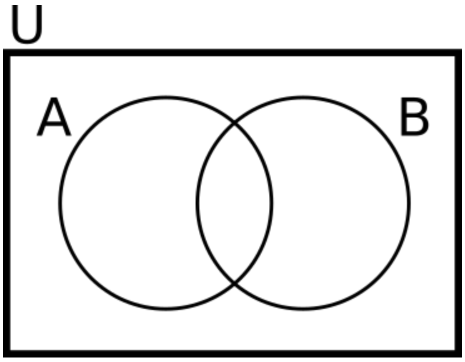
$$P(B_1 \vee B_2 \vee B_3 \vee B_4 \vee B_5 \vee B_6) = P(B_1) + P(B_2) + P(B_3) + P(B_4) + P(B_5) + P(B_6) = 6 \times \frac{1}{36} = \frac{6}{36}.$$

This is the **addition law of probability for mutually exclusive events**.

Laws of Multiplication and Addition for Independent Events

Consider a single trial consisting of two events/statements A and B. We are interested in the compound events/statements $A \wedge B$ and $A \vee B$. This can be depicted in the following **Venn diagram**.

The central region, common to both A and B, is where A and B are both true simultaneously. The two circular regions represent $P(A)$ and $P(B)$, respectively.



Let us consider these two events to occur in the sequence 'A followed by B'. In the first event, suppose that A is true. This covers the left-most circular region. Now, in some of those occasions when A is true, the second event, B, will also be true and these occurrences would be in the central area. This central area represents the probability $P(A \wedge B)$, i.e., the probability that 'A and B are both true'.

When the two events, A and B, are **independent**, then

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

This is what we used in the earlier examples when the vents were independent. The three shaded regions combined represent $P(A \vee B)$. So it can be seen that $P(A)$ and $P(B)$ both include $P(A \wedge B)$. Hence

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$

This is the **law of addition for probability**. In cases where A and B are mutually exclusive, we have $P(A \wedge B) = 0$, because A and B cannot be true simultaneously, and then this law reduces to

$$P(A \wedge B) = P(A) + P(B).$$

This is what we have already used for mutually exclusive events.

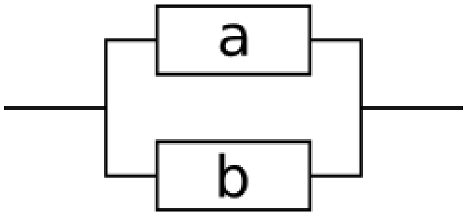
Example 4 A normal pack of 52 playing cards has 26 red cards and 26 black cards. One card is drawn from the full pack at random then replaced in the pack. Then a second card is drawn from the pack. Calculate the probability that both cards are red. Let $A =$ 'The first card is red' and $B =$ 'The second card is red'. Here the two events are **independent** and the probabilities are each $\frac{26}{52} = \frac{1}{2}$. Hence

$$P(A \wedge B) = P(A) \times P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Example 5 Consider the experiment of Example 4 (selection **with** replacement). Here we wish to calculate the probability that at least one of the cards selected is red. This involves statement $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$. Hence

$$P(A \vee B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Example 6 A short network link consists of two sections, A and B, in parallel as shown below.



The link functions if either a or b both function. The two sections are independent in that the functionality of one does not affect the functionality of the other. The probability that a continues to function during a particular time period is 0.9. The corresponding probability for b is 0.8. Consider the following statements:

- $A =$ 'a continues to function during a particular time period.'
- $B =$ 'b continues to function during a particular time period.'

Since a and b are in parallel, the link works if a OR b works, i.e., $A \vee B$.

The probability that the link continues to function during a particular time period is $P(A \vee B)$:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$

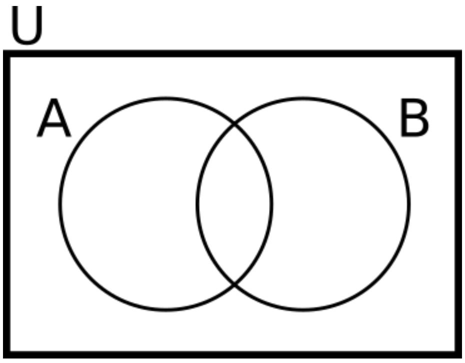
Since A and B are independent: $P(A \wedge B) = P(A) \times P(B) = 0.9 \times 0.8 = 0.72$. So

$$P(A \vee B) = 0.9 + 0.8 - 0.72 = 0.98.$$

This means that there is a 98% probability that the link will function.

Conditional Probability

When the two simple events are not independent, then the process is not quite as simple. Consider a single trial consisting of two events/statements A and B. We are interested in the compound event/statement $A \wedge B$ which can be depicted as in the following Venn diagram. As stated before, the two circular regions represent $P(A)$ and $P(B)$, respectively.



The value of the **conditional probability** $P(A|B)$ (i.e., the probability that B will be true given that A is already true) can be determined by means of our earlier simple definition of probability.

Let there be N outcomes altogether, i.e., number of member of the relevant universal set. Statement A will be true on $NP(A)$ occasions. Of these, B will be true on $NP(A \wedge B)$ occasions. Hence the probability that B will be true, given that A is already true, is

$$P(B|A) = \frac{NP(A \wedge B)}{NP(A)} = \frac{P(A \wedge B)}{P(A)}.$$

This can now be recast as

$$P(A \wedge B) = P(A) \times P(B|A).$$

This is the **law of multiplication** for probabilities of dependent events.

When the two events, A and B, are independent, then $P(B|A) = P(B)$ and the law reduces to

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

This is what we used in earlier examples when the events were independent.

If we considered B to be the first event and A to be the second event, then we would get

$$P(B \wedge A) = P(B) \times P(A|B)$$

which is the same as before but with A and B interchanged.

Example 6 A normal pack of 52 playing cards has 26 red cards and 26 black cards. One card is drawn from the full pack at random and kept. A second card is drawn from the remainder of the pack.

1. Calculate the probability that both cards are red.
2. Calculate the probability that at least one of the cards is red.

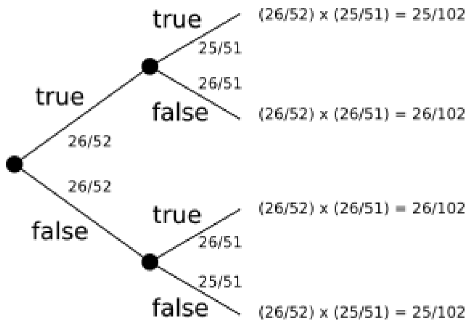
Let A = 'The first card is red' and B = 'The second card is red'.

In the first event, there are 26 outcomes, out of 52, which make A true. Hence $P(A) = \frac{26}{52} = \frac{1}{2}$. In the second event, given that A is true, there are 25 outcomes, out of 51, which make B true. Hence $P(B|A) = \frac{25}{51}$. Consequently,

$$P(A \wedge B) = P(B|A) \times P(A) = \frac{25}{51} \times \frac{1}{2} = \frac{25}{102}.$$

Hence the probability that both cards are red is $\frac{25}{102}$.

Of course, there are other outcomes. Statement A could be false, as could B. All of the outcomes, with respect to our statements A and B can be shown, along with the corresponding probabilities, in a probability tree, as follows.



The probabilities associated with each of the simple events, A and B, are obtained by considering the number of ways of obtaining 'success' out of the total number of possibilities.

We did this earlier in Example 5. As a second illustration of this idea, suppose that A is false. There are 51 cards left (independent of the outcome of A) of which 26 yield a red card and so probability that B is true given that A is already false, or simply $P(B|\neg A) = \frac{26}{51}$.

Such a probability tree is very useful but is practicable only when there is just a small number of branches emanating from each node. In a case with many possible events, we might tabulate the possibilities as we did with the dice example in Example 1.

For the second calculating (that at least one of the cards is red) we can rearrange our formula for probabilities:

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$$

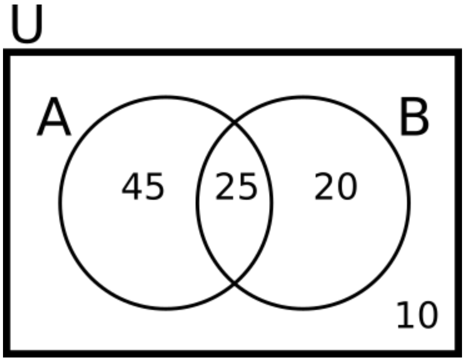
In this case we are now calculating whether the first card is red, the second card is red, or both cards are red, i.e., at least one of the cards is red. So

$$P(A \vee B) = \frac{26}{52} + (\frac{25}{102} + \frac{26}{102}) - \frac{25}{102} = \frac{77}{102}.$$

Example 6 The following Venn diagram represents the occurrence of two events A and B. The outer rectangle is the universal set and the two circular regions represent A and B as shown. The values in the diagram indicate the number of outcomes (total 100) in each case.

We want to determine $P(A)$ and $P(B)$:

- $P(A) = \frac{45+25}{100} = \frac{70}{100} = 0.7$,
- $P(B) = \frac{25+20}{100} = \frac{45}{100} = 0.45$.



Now suppose that we want to determine $P(B|A)$ and $P(A|B)$. The probability $P(B|A)$ means that B occurs given that A has already occurred. The number of outcomes where A has occurred is $45+25=70$. Out of these 70 outcomes, B occurs in 25 cases and B does not occur in 45 cases. So $P(B|A) = \frac{25}{70} = 0.357$ (3 d.p.). Alternatively, we can use the formula

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

From the diagram, $P(A \wedge B) = \frac{25}{100} = 0.25$, so $P(B|A) = \frac{0.25}{0.7} = 0.357$ (3 d.p.). Similarly, $P(A|B) = \frac{25}{45} = 0.556$ (3 d.p.).

Clearly $P(B)$ and $P(B|A)$ are different, just as $P(A)$ and $P(A|B)$ are different. This means that A And B are **not independent**.