Probability

Contents

2 Probability – Foundations

Probability is the area of mathematics which deals with non-deterministic phenomena or occurrences, usually described mathematically as events. Non-deterministic events are those which cannot be predicated with certainty.

There are several ideas central to the topic, these will be defined first; everything we come to study in this area, including statistical testing and statistical inference, can only be properly understood if these fundamental definitions are kept in mind.

2.1 First Definition of Probability

We will start by defining events, experiments and probability. We will introduce these ideas, with the example of the dice, in an intuitive framework and then move on to the formal mathematical definitions.

2.1.1 Definition: An Event

An *event* is a well-defined occurrence that may or may not happen.

A very important part of this definition is the idea that the event is *well-defined*. An event must be exactly described so that we know whether or not it has happened. This definition of an event also leads to an exact definition of its probability. We will define this idea in two steps, the first giving a general idea of probability, the second giving a quantifiable version.

2.1.2 First Definition: Probability

The *probability* of an event is a measure of how likely it is to occur. It is a number between 0 and 1, where a probability of 0 for an event means it will not occur and a probability of 1 means it definitely will occur. The higher the probability of an event, the more likely it is to occur.

Any introduction to probability could quickly become mired in difficult philosophical debates, but this concept of probability as a measure of how confident we are a non-deterministic event will occur is common to all.

The word 'probability' derives ultimately from the Latin *probabilitas*, a measure of the authority of a witness in a legal case in mediaeval Europe (this is often the same as the witness's nobility); the word *probity* comes from the same root. It could be argued that this differs considerably from the modern meaning of probability, a measure of the weight of empirical evidence, arrived at from inductive reasoning and statistical inference, but the idea of *confidence* is still there.

2.1.3 Notation for probability

Mathematically, an event can be denoted by any symbol. Thus the event E could be the event of a particular number showing up on a dice. With this in mind, the probability of event E is written mathematically as

$$P[E].$$

Note the use of a functional-style notation; $P[.]$ reminds us that the probability is a function of the event and indeed will have certain properties expressed in functional notation.

The symbol E may be replaced by a description of the event if it is mathematical in nature, thus if X is a variable to be measured we may attempt to find the probability it produces a value above 4:

$$P[X > 4].$$

More of this in due course. Here is a simple example of the notation.

"Met Eireann has stated that there is a 35%
chance of rain in Dublin today."

Set R to be the event that there is rain in Dublin today, however it is actually defined. The casual '35% chance' becomes the probability of 0.35 and we write:

$$P[R] = 0.35.$$

2.2 Second Definition of Probability

The first definition given for the probability of an event gave a number as a measure of the confidence it will occur. This number may be defined if the event may occur many times. For example, the probability of a given face coming up on a dice may be defined as 1 divided by 6, if it is assumed each face comes up an equal number of times when the dice is thrown a large number of times. To allow us to follow this line of thought for the definition of probability, we need another concept, that of an experiment.

2.2.1 Definition: Experiment

An action, with a well-defined set of outcomes, that can be repeated a large number of times, is called an *experiment*.

The importance of this idea of an experiment is that an event can now be regarded as the result of an experiment and defined in terms of one or more of the outcomes of the experiment.

In the case of the dice, throwing the dice is the experiment, with the numbers which come up as the outcomes. The events such as ‘getting an even number’ or ‘getting a 5’ are now defined as made up of the outcomes.

We will now give a definition of probability with these concepts which will give us a number.

2.2.2 Second Definition: Probability

Let E denote an event, which is the possible outcome of an experiment. Let N be the number of times the experiment is repeated. Let N_E be the number of times event E turns up. Then the probability of E occurring may be defined as the proportion

$$P[E] = \frac{N_E}{N},$$

evaluated as the number N gets bigger. This is then the limit:

$$P[E] = \lim_{N \to \infty} \frac{N_E}{N}.$$

The probability $P[E]$ is therefore the proportion of times the event E occurs, as the experiment which leads to its occurrence is repeated an ever larger number of times. It is called the *frequentist* interpretation of probability and is regarded as an objective, empirical and data-driven interpretation.

The is a hidden assumption in this definition, that the limit

$$P[E] = \lim_{N \to \infty} \frac{N_E}{N}.$$

does indeed converge. The Law of Large Numbers is a statement that this will happen; it can alternatively be proved or regarded as an axiom to establish probability science.

The previous, confidence based, definition is not inconsistent with the frequentist definition, but may lead in different directions. Bayes theorem, which we will meet in due course, uses the first definition to allow for an increasingly correct figure to be given for a probability based on successively acquired information.

2.2.3 The Single Dice as an Experiment

In the example of the throw of one dice, when the experiment is done, that is, the dice has been thrown, we are interested in the number left facing up. The set of possible outcomes is

$$\{1, 2, 3, 4, 5, 6\}$$

The various possible events can be defined in terms of these outcomes. To see what the probabilities are, the dice can be thrown many times, and the outcomes, and so events, counted. It would be expected that after a large number of throws, each side of a fair dice would come up a roughly equal number of times. Indeed, this is the definition of a fair dice. Therefore, if the dice is thrown N times, where N is a large number, it follows that if each face comes up approximately the same number of times n, and

$$6n \approx N.$$

From the second, quantitative definition of probability, the proportion of times each number comes up is approximately

$$\frac{n}{N} \cong \frac{1}{6}.$$

As the number of throws N increases without limit, the ratio becomes closer to the value 1/6. This means we can define the probability as follows: if E is the event of a given number coming up when the dice is thrown, then

$$P[E] = \frac{1}{6}.$$

2.2.4 Empirical Data and Probabilities

The probability in the previous example came up as a result of theoretical reasoning. Now let is look at some empirical examples, where we rely on evidence in the form of data. Pay heed to the use of the word *random* – the last proportion estimated is based on a randomly selected group of subjects, where there is no reason to think they are any different from the overall population. More fundamentally, this means that we have no reason to suppose that the proportion we calculate from the sample will change.

Empirical Data	***Probability***
Clinical data suggests that 8% of the population born in Ireland has O Negative blood-type.	The probability that a person chosen at random and born in Ireland has this blood-type is 0.08.
A recent census established that 6.2% of the population of Ireland have no religion.	The probability that a person in Ireland, chosen at random, has no religion is 0.062.

A company manufactures a certain type of component, by the thousands. A random sample of 200 components is taken and 5 are found to be faulty.	The company management concludes that the probability of an individual component being faulty is 0.025.

2.2.5 Rolling Two Dice

We will now look at this definition of a probability with the example of the throw of two dice. For this case, we will calculate the probability of the result adding up to 5.

Firstly, let E be the event that we roll two dice and we get a sum of five. Theoretically, to find the probability, we need the number N, the number of times the experiment is repeated, and N_E, the number of times the event E turns up. Then the probability of E occurring, is the fraction

$$P[E] = \frac{N_E}{N}$$, evaluated as the number N gets bigger.

However, similar information is available to us for the individual dice, which we will now use. Firstly, note there are 36 possible outcomes in all, each pair of numbers. These outcomes are are all equally likely to come up as the experiment is repeated a large number of times. Only 4 have the sum of 5:

(1, 4), (2, 3), (3, 2) and (4, 1)

This list of outcomes is the event ‘the sum equals 5’.

If the dice are thrown a large number of times, then each pair comes up 1/36 of the time, and the pairs adding to 5 come up 4/36 of the time. As the number of rolls increases, we expect this proportion to hold; a non-trivial point mentioned above as the law of large numbers. The proportion of times the sum of 5 comes up is then

$$\frac{N_E}{N} = \frac{4}{36}, \text{ so } P[E] = \frac{N_E}{N} = \frac{4}{36} = \frac{1}{9}.$$

As this example shows, the probability of a particular event can often be found by considering all the possible outcomes of the experiment being run, all of which are equally likely, and using a counting process with this information to calculate

$$P[E] = \lim_{N \to \infty} \frac{N_E}{N}.$$

2.2.6 The Sample Space

In the previous example, there was an important distinction between the outcome of the experiment, the pair of numbers which turn up, and the event, which is that they add up to 5. More specifically, the outcomes were the pairs

(1,1), (1,2), …, (1,6), (2,1), (2,2),…, (6,1), (6,2),…, (6,6).

The event was the list of 4 outcomes:

(1, 4), (2, 3), (3, 2) and (4, 1).

Therefore we see clearly that the event is a subset of this list of outcomes.

2.2.7 Definition of Sample Space

The sample space is the set of all possible outcomes of an experiment. Any event which arises as a result of that experiment is a subset of the sample space.

2.2.8 An Example of the Sample Space

For the case of two dice, we will calculate the probability that the sum is below 6. Once again we need to find the ratio N_E/N, for very large numbers of repetitions of the experiment. The most methodical way of counting the number of outcomes that satisfy the definition of the event is to calculate the sum for each possible outcome. Here it is:

	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

This is equivalent to the sample space, since every possible event is a subset of this table. It also provides the sum, so we can check whether a given outcome is in the event subset or not.

To find the probability of this event, count those outcomes where the sum is below 6; this is 10, so the probability is 10/36.

The full information is given by listing each outcome and the corresponding value of the sum which allows us to correctly count the outcomes that make the event. This might look like:

((1, 1), 2), ((1, 2), 3), …, ((1,6),7), ((2,1), 3), ((2,2), 4), …, ((6,1), 7), ((6,2),8) …, ((6,6), 12).

We are now getting on to the idea of a random variable, but for the time being the third number, in this specific case the 'sum', is the criteria by which we determine whether the event has happened or not.

2.2.9 The Difference

In the same experiment, rolling two dice, we calculate the probability that the difference of the two numbers is 2. This calculation is done by drawing up the sample space for the experiment and for 'the difference'. Then find the probability by counting, as before. This is essentially the same list of outcomes, but now with an alternative label, the difference, which is drawn from the list of numbers:

$$\{0, 1, 2, 3, 4, 5\}.$$

Here is the event space for ‘the difference’:

	1	*2*	*3*	*4*	*5*	*6*
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

There are 8 results where the difference is 2, so the probability is:

$$8/36 = 2/9.$$

In the case of rolling two dice and noting the sum,

- The experiment was rolling the two dice.
- The sample space was the set of all pairs of numbers, with the sum/difference noted. These are the outcomes of the experiment.

2.2.10 Complementary Events

Now calculate the probability that the sum is equal to or greater than 6. The event we are investigating here is the exact opposite of the previous event, the sum less than 6. From our existing information, the number of outcomes equal to or greater than 6 is

$$36 – 10 = 26.$$

The probability of the sum being equal to or greater than 6 is then 26/36.

We were able to do this subtraction with the aid of the sample space and it leads us to a particularly important rule.

In an experiment with a number of possible outcomes, let A be an event and let B be the exactly opposite event. The event B is called the complement of A. Then

$$P[A] + P[B] = 1.$$

In other words, when two distinct events between them cover all possibilities eventualities, the sum of their probabilities is 1.

2.2.11 The Birthday Enigma

Here is an example which gives what seems like an unusual result, using the same ratio of outcomes from a sample space.

"In a group of 25 people, calculate the chances that two or more share the same birth date."

This event is more likely than not; consider a slightly more abstract, but simpler, version of the same question.

"20 people are asked to choose a number from 1 to 100. Calculate the probability that two or more people pick the same number."

It is simpler to look at the converse question and find the probability that no two people pick the same number. Let S be the event that two or more of the numbers chosen are the same. Let T be the complementary event that none of the numbers are the same.

The probability $P[T]$ can be found by seeing how many ways in which the 20 people can pick their numbers with no restrictions, the quantity N. Then find the number of outcomes which satisfy the criteria for the event, the quantity N_E. We are implicitly assuming that each outcome is equally likely to come up. The probability is then the ratio

$$P[T] = \frac{N_T}{N}.$$

The number of unrestricted possible combinations of the 20 numbers is:

$$100 \text{ x } 100 \text{ x } \ldots \text{ x } 100 = 100^{20} = 10^{40}.$$

For the number of ways numbers can be chosen so none are the same:

- The first person has a choice of the full 100.
- To ensure the first number is not picked again, the second person has the choice of the other 99 numbers.
- To ensure the first or second number is not picked again, the third person has the choice of the other 98 numbers.
- Continuing like this, the last person will have a choice of 81 numbers, so the full number of combinations will be 100 x 99 x 98 x …x 81

This number is

$$100 \text{ x } 99 \text{ x } 98 \text{ x } \ldots \text{x } 81 = {}^{100}P_{20} = 1.3\text{x}10^{39}.$$

If the 20 numbers are chosen at random, the probability that none are the same is

$$P[T] = \frac{{}^{100}P_{20}}{100^{20}} = 0.13$$

and so $P[S] = 1 - 0.13 = 0.87$.

2.2.12 Birthdays

Now return to the original question of birthdays. In a group of 25, calculate the probability that two or more people share a birthday.

Let S be the event that two or more of the people share the same birthday, and let N be the complementary event that none are the same. Then the probability of nobody having the same birthday among this group is

$$P[N] = \frac{{}^{365}P_{25}}{365^{25}}.$$

The figures are:

$$365^{25} = 1.14\text{x}10^{64}, \text{ and } {}^{365}P_{25} = 4.92\text{x}10^{63}.$$

The probability of nobody having a birthday in common is then

$$P[N] = 0.43,$$

So

$$P[S] = 0.57.$$

Therefore in a group of 25 people, the probability that two or more people share the same birthday is 0.57.

2.2.13 Definition – Mutually Exclusive Events

Two events A and B, the possible results of the same experiment, are said to be *mutually exclusive* if it is impossible for them to happen together. If two events A and B are not mutually exclusive, that is, A and B can occur together, they are said to be mutually *non-exclusive* events.

This allows us to discuss the following scenario.

2.2.14 Three or More Complementary Events

In the case of throwing two dice and noting the sum, we calculate the probability of the following events; the sum is less than 7, the sum is equal to 7, and then the sum is greater than 7. Counting the outcomes which give the correct respective sum, there are

15, 6, and 15 possibilities respectively.

The probabilities are then

15/36, 6/36 and 15/36.

Note that these all add up to 1, as they should from the counting process. A generalisation of the rule for complementary events can now be based on this example of the dice.

Say an experiment has n mutually exclusive possible events:

$$E_1, E_2, E_3, \ldots, E_n,$$

and no others. Then

$$P[E_1] + P[E_2] + \ldots + P[E_n] = 1.$$

In other words, when all distinct possibilities have been taken care of, the sum of the probabilities is 1.

2.3 The Laws of Probability

In this section we will look at motivating examples of ways in which probabilities involving two or more events can be calculated, depending on how the events are related to one another. We will then rephrase these laws as the fundamental axioms of probability using set theory.

We have already seen the rule for complementary events. If A is an event and B is the exact opposite event, then

$$P[A] + P[B] = 1.$$

In other words, the probabilities of two distinct events, which cover all possibilities, add up to 1.

A generalisation of the above rule concerns n distinct possible events:

$$E_1, E_2, E_3, \ldots, E_n,$$

and no others. Then

$$P[E_1] + P[E_2] + \ldots + P[E_n] = 1.$$

In other words, when all distinct possibilities have been taken care of, the sum of the probabilities of the events is 1.

The 'addition law' concerns how we calculate the probability of the event 'A or B'.

2.3.1 The Addition Law

If two events A and B are mutually exclusive, then the following law holds:

$$P[A \text{ or } B] = P[A] + P[B].$$

If the two events are not mutually exclusive, the law above is broadened so that the probability of A or B occurring is given by:

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B].$$

This means the probability of one of event A or event B occurring is given by the sum of their two probabilities. If the events are not mutually exclusive, the probability of one of event A or event B occurring is given by the sum of their two probabilities, less the probability of them both occurring.

2.3.2 An Example of the Addition Law

If a single dice is thrown, we determine the probability of getting a multiple of 2 or a multiple of 3, and then the probability of one or the other. Let A be the event of getting a multiple of 2, and B be the event of getting a multiple of 3.

The probability of scoring a multiple of 2 is that of getting a 2, 4 or 6 so that

$$P[A] = 3/6 = ½.$$

The probability of scoring a multiple of 3 is that of getting a 3 or a 6, so that:

$$P[B] = 2/6 = 1/3.$$

Now both of these events could happen at the same time; a number can be a multiple of 2 and 3, and one of these is on the die: 6. It then follows that A and B are non-exclusive events, so

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B].$$

The event 'A and B' means that a number is a multiple of 2 and 3, so this is equivalent to getting a 6. Thus

$$P[A \text{ and } B] = 1/6.$$

The final probability is then

$$P[A \text{ or } B] = 2/6 + 3/6 - 1/6 = 4/6 = 2/3.$$

In this case, finding the probability $P[A \text{ or } B]$ from scratch would not require too much calculation. It would simply be a matter of counting the number of ways a number on the dice could be a multiple of 3 *or* a multiple of 2. The ways of getting this are 3 and 6 for the first, and 2, 4 and 6 for the second. This is a list of 4 distinct numbers, so the answer is 4/6 = 2/3. This confirms the answer already found and is also 'computationally' the same.

2.4 Independent and Dependent Events

Consider the rolling of a dice on two occasions. The outcome of the first throw will not affect the probabilities for the second throw; the dice is picked up and thrown again with no link to the last throw. In this context we say that each throw of the dice has *no memory*.

Now consider the event of drawing a lemon from a bag of lemons and limes without replacing it, and the second event of drawing a lime after this. Clearly whether the first event *does* or *does not* happen, this will alter the probability of the second event. These two events are therefore not independent, they are dependent. We will see the details of this soon.

2.4.1 Definition – Independent Events

Two events are *independent* when the occurrence of one event does not affect the probability of the occurrence of the second event. If the outcome of one event does affect the probability of the second event, they are said to be *dependent*.

To go back to the case of two dependent events, consider the experiment of picking fruit from a bag of 5 lemons and 10 limes. Let A be the event of picking a lemon from the bag, without replacement, and let B be the event of picking a lime after this. The probability of A is:

$$P[A] = 5/15 = 1/3.$$

The fruit is not replaced, and the same experiment is carried out again. If the first fruit *was* a lemon, there are 4 lemons among the 14 remaining fruit, so then the probability of getting a lime the second time is:

$$P[B] = 10/14.$$

If the first fruit *was not* a lemon, there are 5 lemons in the 14 remaining fruit. Therefore the probability is:

$$P[B] = 9/14.$$

Thus the outcome of the first experiment, drawing a lemon or not, has affected the probabilities for the second. The two events are not independent, in other words they are dependent.

2.4.2 The Multiplication Law of Probability (1)

For two *independent events* A and B, the probability of the occurrence of both events A and B, is given by

$$P[A \text{ and } B] = P[A]P[B]$$

The probability of the occurrence of both events is the product of the two individual probabilities.

2.4.3 An Example of Independent Events

A single fair dice is thrown 4 times. Calculate the probability of getting the same number four times in a row.

The event of rolling the die each time is an independent event. Each time, the probability of getting a specific number is 1/6. Thus using the multiplicative law, the probability is

$$(1/6)^4 = 1/6^4.$$

For the second problem, let A_1 be the event of getting four 1's in a row, A_2 that for four 2's in a row and so on. Then the second event of the same number four times in a row is

$$A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_6.$$

Each of these events has the same probability, so therefore

$$P[A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_6] = P[A_1] + P[A_2] + \dots + P[A_6] =$$
$$= 6\text{x}(1/6)^4 = 1/6^3.$$

2.4.4 The Binomial Distribution

Given it is known that 11% of the Irish population is left-handed, find the probability of finding 3 people who are left-handed, in a randomly selected group of 12 Irish people.

Look at one case of the first 3 of the 12 being left-handed.

If the group have been randomly selected, we may assume that the probability that each individual being left-handed is 0.11.

We can also assume there is no connection between them, so each person being left or right-handed is an independent event, so the probability of the first 3 being left-handed, and therefore the remaining 9 being right-handed, is

$$(0.11)^3.(0.89)^9.$$

Any selection of 3 from the 12 could be the relevant left-handed people and the probability of that selection is the same since multiplication is commutative.

These possible outcomes are all mutually exclusive, so to find the overall probability, add on the probability

$$(0.11)^3.(0.89)^9$$

for each combination, in other words, multiply by ${}^{12}C_3$. The probability of 3 left-handed people in a group of 12 is therefore

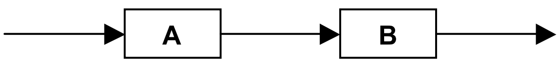
$${}^{12}C_3(0.11)^3.(0.89)^9.$$

This is an example of what is called the Binomial distribution.

2.5 Some Examples of Complex Systems

In each of the following systems, the probability that each individual component of type A will fail is 0.03, and the probability that each individual component of type B will fail is 0.05. All components A and components B are independent. For each system, we will calculate the probability that the system overall will work.

2.5.1 Example 1

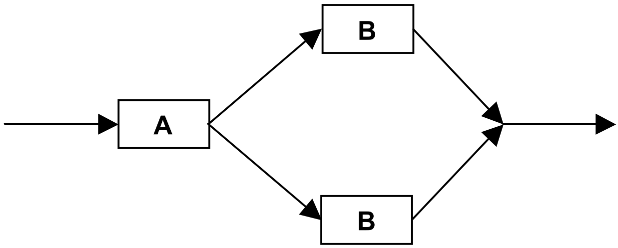


For this example, clearly both components have to work so that the system overall works. The two events, *A* and *B*, are independent, so the probability they both work is:

$$0.97x0.95 = 0.9215.$$

This then is the probability the system works.

2.5.2 Example 2



In this case, going from left to right, the first component A has to work, followed by the second stage, which is *either* one B or the other. For the system overall, both stages must work. Therefore:

$$P[\text{system works}] = P[\text{stage 1 works and stage 2 works}] =$$
$$= P[\text{A}] \text{ x } P[\text{stage 2 works}].$$
$$= P[\text{A}] \text{ x } P[\text{B or B}].$$

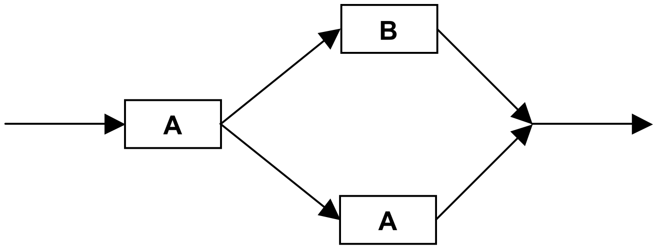
In the second stage, either *B* working are not mutually exclusive. So use the second version of the addition law, this is

$$P[\text{B or B}] = P[\text{B}] + P[\text{B}] - P[\text{B}]P[\text{B}].$$

The laws then mean that the probability the whole system works is

$$P[\text{system works}] = P[\text{A}] \; (\; P[\text{B}] + P[\text{B}] - P[\text{B}]P[\text{B}] \;) =$$
$$= 0.97\text{x}0.9975 = 0.9676.$$

2.5.3 Example 3



In this case, going from left to right, the first component A has to work, followed by the second stage, which is *either* the B or the second A.

For the system overall, both stages must work.

$$P[\text{system works}] = P[\text{stage 1 works}] \text{ x } P[\text{stage 2 works}]$$

$$= P[\text{A}] \text{ x } P[\text{A or B works}]$$

So to find the probability this part of the system works, we have to use the second version of the addition law, this is just

$$P[\text{A or B}] = P[\text{A}] + P[\text{B}] - P[\text{A}]P[\text{B}].$$

The laws then mean that the probability the whole system works is

$$P[\text{system works}] = 0.97\text{x}(0.97 + 0.95 - 0.97\text{x}0.95) = 0.969.$$

2.6 Conditional Probability and Bayes Rule

To deal with dependent events, some notation will be needed to indicate when the probability of one event depends on whether another event has happened.

This is the situation of two events A and B, where the occurrence or not of event A does effect the probability of event B. In other words, B is dependent on A.

2.6.1 Notation – Conditional Probability

Let A and B be two events, where event B is dependent on event A. The probability of event B, given that event A has already occurred, is denoted by:

$$P[B \mid A].$$

This is 'the probability of B, given A.'

For two independent events A and B, by definition, the fact that A has already occurred does not affect the probability of event B. It then follows that

$$P[B \mid A] = P[B].$$

If the events are dependent, these probabilities are not the same.

2.6.2 The Multiplication Law of Probability (2)

Now consider two events A and B; with event B dependent on A. The probability of the occurrence of both events is given by

$$P[A \text{ and } B] = P[A]\, P[B \mid A]$$

The probability of both events happening is the probability of A times the probability of B, given that A has occurred.

2.6.3 An Example of Conditional Probabilities

A box contains five 10kΩ resistors and twelve 20 kΩ resistors. Determine the probability of randomly picking a 10kΩ resistor from the box and then a 20 kΩ resistor.

Let event A denote the event of picking a 10 kΩ resistor, and let B denote the event of picking a 20 kΩ resistor.

The first probability is just $P[A]$, and since the total number of resistors is 17, it is

$$P[A] = 5/17.$$

To find the probability of both, that is, $P[A \text{ and } B]$, observe that B depends on A, since A is the event that is happening first. The probability law for dependent events must be used:

$$P[A \text{ and } B] = P[B \mid A]\, P[A].$$

To find $P[B \mid A]$, we need the probability that a second resistor picked from the box will be a 20kΩ resistor, providing that the first one was a 10kΩ resistor. For this case,

$$P[B \mid A] = 12/16 = 3/4.$$

So the probability of both events, picking a 20Ω k resistor after getting a 10 kΩ resistor is:

$$P[A \text{ and } B] = P[B \mid A].P[A] = 3/4 \text{ x } 5/17 = 15/68.$$

2.6.4 Bayes Rule

The second multiplication law, dealing with conditional probabilities, leads to a deceptively powerful result in probability known as Bayes Rule or sometimes Bayes Theorem. Recall that

$$P[A \text{ and } B] = P[A]\, P[B \mid A]$$

We reorganise this equation as

$$P[B \mid A] = \frac{P[B \cup A]}{P[A]}.$$

The 'U' symbol is used for the concept of 'and' here, a usage which will become more useful soon.

This equation is sometimes given as the definition of conditional probability, but for the time being we will treat it as a re-statement of the multiplicative rule above. It has powerful implications as the following example illustrates.

2.6.5 An Example of Bayes rule

Consider the case of a patient undergoing a medical check for a disease. Let T be the event the test gives a positive result, let D be the event the patient has the disease. Let us firstly identify the known probabilities and comment on what they are likely to be.

The probability of the disease occurring at random in a patient is known; $P[D]$. It is typically rare.

The probability of getting a positive result given a patient has the disease is known; $P[T|D]$. It is typically high.

The probability that the patience gives a positive result when they do not have the disease is $P[T|D']$, it is typically low.

The second and third probabilities could be found by simply noting all previous patients who took the test and seeing whether, if untreated within a given time period where they were not treated, they developed the symptoms of the disease.

In the event of a positive result, this data is often assumed to be the only information required; the test is good therefore if the patient gets a positive result the patient and doctor are worried. However, what is required is the probability the patient has the disease given a positive result, which is $P[D|T]$, and is given by Bayes rule:

$$P[D \mid T] = \frac{P[D \cup T]}{P[T]}.$$

The probability of the event $D \cup T$ can be calculated by a second application of the multiplicative law. The same goes for the probability T.

Firstly, the probability $P[D \cup T]$ is given by

$$P[D \cup T] = P[T|D]\, P[D].$$

These probabilities are all known from the data above. The probability $P[T]$ is given by

$$P[T] = P[D \cup T] + P[D' \cup T] = P[T|D]\, P[D] + P[T|D']\, P[D'].$$

Again, these probabilities are either known or can be calculated from the data given. In fact what we have is:

$$P[D \mid T] = \frac{P[T \mid D]P[D]}{P[T \mid D]P[D] + P[T \mid D']P[D']}.$$

This version of the equation clearly shows the way in which the rule gives us the probability $P[D|T]$ from the probability $P[T|D]$.

To see how dramatic the results of the use of Bayes rule in this context can be, let us use some data. For the disease and the efficacy of the testing procedure, let the data be as shown:

- The probability of the disease occurring in a patient at random is $P[D] = 0.001$. This is quite rare.
- The probability of getting a positive result given a patient has the disease is $P[T|D]$ is 0.99. The test is very good.
- The probability a patience gives a positive result when they do not have the disease is low, $P[T|D'] = 0.02$.

All these results, as noted above, would lead to great concern in the event of a positive result. However, let us calculate the relevant probability $P[D|T]$ with Bayes rule:

$$P[D \mid T] = \frac{P[T \mid D]P[D]}{P[T \mid D]P[D] + P[T \mid D']P[D']}.$$

Looking at the top line we have all the probabilities in $P[T|D]$ $P[D]$, then in the bottom line we just need:

$$P[D'] = 1 - P[D] = 0.999.$$

Therefore

$$P[D \mid T] = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.02 \times 0.999}.$$

This is 0.0472, a strikingly small probability. To gain some insight into why results like this may happen, rearrange the equation to give:

$$P[D \mid T] = \frac{1}{1 + \frac{P[T \mid D']}{P[T \mid D]} \times \frac{P[D']}{P[D]}}.$$

The probabilities in the ratios below the line are:

$$\frac{P[T \mid D']}{P[T \mid D]},$$

the ratio of getting a positive test result when the patient does not or does have the disease, and

$$\frac{P[D']}{P[D]},$$

the odds of not having the disease. In this case, the first ratio was quite small, as would be expected, but the second ratio is a very large number, which reduced the overall probability $P[D|T]$. This suggests another formulation of Bayes rule using odds.

2.6.6 Odds form of Bayes Rule

We are familiar with the concept of the odds of an event happening, so for an event E then the odds for E is the probability it does happen divided by the probability is does not.

$$O[E] = \frac{P[E]}{P[E']}, \text{ where } P[E'] \neq 0.$$

By rewriting the current form of Bayes rule

$$P[A \mid B] = \frac{P[B \mid A]P[A]}{P[B \mid A]P[A] + P[B \mid A']P[A']},$$

for the event A' and then dividing the two identities we find the following relation:

$$\frac{P[A \mid B]}{P[A' \mid B]} = \frac{P[B \mid A]P[A]}{P[B \mid A']P[A']},$$

or in terms of odds,

$$O[A \mid B] = \frac{P[B \mid A]}{P[B \mid A']} O[A].$$

This equation now states that the odds of event A happening, given B, is the original odds of event A happening times the ratio

$$\frac{P[B \mid A]}{P[B \mid A']}.$$

This quantity is known as the relative likelihood of B with respect to A. Essentially we have an equation that starts with the odds of an event A, then updates these odds based on the information from B.

The most significant way of regarding the meaning of Bayes rule, especially in this form, is that it is in effect updating a probability, in the form of odds, with new information. Consider the following example.

2.6.7 Example – Crime and Punishment

A crime has occurred and only one of two individuals A and B could have committed it. Two pieces of additional evidence have emerged; the perpetrator was of blood type O neg and so is person A. nothing is known about the blood type of B. The proportion of the population who are O neg is 0.08. We will see how this changes the odds on X's possible guilt.

In this context, we refer to the probabilities or odds before the new evidence as the *a priori* odds and those after the new evidence as the *posterior* odds.

Let H be the hypothesis that person X is the perpetrator. Before the new evidence, the odds of this are 1. Let E be the event of the evidence appearing. The relative likelihood of H given E is calculated first.

The probability of the blood found being O neg if X committed the crime is 1; $P[E|H] = 1$. If X did not commit the crime, then the probability of the blood being O neg is the same as for the general population; $P[E|H'] = 0.08$. The likelihood ratio is therefore 12.5.

The calculation with the Bayes rule then says

$$O[H \mid E] = \frac{P[B \mid A]}{P[B \mid A']} O[H] = 12.5 \text{ x } 1 = 12.5.$$

Therefore the odds of X being guilty have increased considerably.

2.7 Set Theory Formulation of Probability

The definitions and laws of probability we have seen so far will be rephrased here as laws on sets.

2.7.1 The Sample Space and the Universal Set

Let an experiment have a sample space *X*, from which the set *U* is constructed, where every event *E* resulting from the experiment is a subset of *U*. Thus for any event we can say that

$$E \subset U\,.$$

The set *U* is called the universal set, or sometimes the event space, as noted already. It is also known as the universal event; it may be viewed as the union of all possible events. In this context, set Ω to be the set

$$\Omega = \{E \mid E \subset U\}.$$

So the set Ω is the set of all subsets of *U*. A probability may now be thought of as a function

$$P\colon \Omega \to [0,1].$$

We will return to this idea to give the function *P* the properties it needs to be a probability as we understand it.

2.7.2 Complementary Events

Let *A* be an event resulting from the same experiment, so $A \subset U$.
Let *B* be the complement of *A*, so

$$B = A' = \{x \mid x \in U, x \notin A\}.$$

then

$$P[A] + P[B] = 1.$$

We can now rewrite the definition of mutually exclusive events.

2.7.3 Definition – Mutually Exclusive Events

Let two events A and B be the possible results of the same experiment. Let U be the corresponding universal set, so that

$$A, B \subset U .$$

The events A and B are said to be mutually exclusive if

$$A \cap B = \{\}.$$

Now say an experiment has n mutually exclusive possible events:

$$E_1, E_2, E_3, \ldots, E_n,$$

and no others, which also satisfy

$$E_1 \cup E_2 \cup \ldots \cup E_n = U ,$$

then

$$P[E_1] + P[E_2] + \ldots + P[E_n] = 1.$$

In other words, when all distinct possibilities have been taken care of, the sum of the probabilities is 1.

2.7.4 Formal Definitions

We will now give more formal definitions of the concepts introduced above, motivated by the examples given and the intuitive ideas of probability.

2.7.5 Definition – Sample Space

Consider an experiment that can produce a number of outcomes. The set of all possible outcomes of the experiment is called the sample space. Any event is a subset of the sample space.

2.7.6 Definition – Event Space / Sigma Algebra

For an experiment that produces outcomes and a sample space, the event space or universal event, is the union of all possible events that can result from the experiment. The sigma algebra is the set of all possible subsets of the sample space.

This object is called an algebra because it behaves like numbers with addition, subtraction etc. or like matrices with their distinct algebra. In this case, the operations will be unions, intersections and other operations from set theory.

These ideas allow us to give a formal definition of a probability. It will be defined as a map from the sigma algebra to the interval [0,1]

$$P\colon \Omega \rightarrow [0,1],$$

with certain properties that come from our understanding of what a probability should be.

2.7.7 Third Definition: Probability

Let an experiment have a sample space X, from which we construct the universal set U. Let Ω be the sigma algebra for the experiment, so

$$\Omega = \{E \mid E \subset U\}.$$

A probability is a function

$$P: \Omega \rightarrow [0,1],$$

with the following properties:

1. Let $E_1, E_2, E_3, \ldots, E_n$, be a list of n mutually exclusive events, then $P\left[\bigcup_i E_i\right] = \sum_i P[E_i]$. In particular, if $\bigcup_i E_i = U$, then $\sum_i P[E_i] = 1$.
2. $P[U] = 1, P[\{\}] = 0$.
3. Let A and B be two events, subsets of U, such that $A \subset B$. Then $P[A] < P[B]$. The more the event includes, the higher the probability. This is called monotonicity.
4. Let A and B be two events, subsets of U. Then $P[A \cup B] = P[A] + P[B] - P[A \cap B]$. This is the addition law; in the simpler case where $A \cap B = \{\}$, we have $P[A \cup B] = P[A] + P[B]$.

This definition of a probability function can be set up with three axioms and then a sequence of theorems following from them; generally property 1 is the most fundamental and properties 2 and 3 follows from it. The well-known rule for complementary events follows directly from property 1. Thus the laws of probability we have seen already follow from this formulation.