Chapter 1: Basic rules

1 Introduction

The objective of this chapter is to develop some of the basic rules of probability theory without engaging in mathematics or philosophy. Strictly, we are establishing laws or theorems rather than rules, but, in our informal context, it seems more appropriate to say rules. We rely on common sense and examples from dice throwing. This brings us to one of the important questions of probability theory: what is the singular of dice, ‘die’ or ‘dice’? In popular use, it is often dice, but we will use the conventional dictionary ‘die.’ A die, plural dice, is a cube; it has six faces, each one an identical square. The six faces are generally numbered 1 to 6, though other arrangements are possible. Each of the six faces are equally likely to land face up when a fair die is thrown. Dice are an essential part of many board games from backgammon to snakes and ladders, monopoly to my own favourite ludo.

2 Outcomes and frequency

The standard, fair die has six faces, numbered 1 to 6. We can imagine the faces of a die being unwrapped from the cube to give the figure below.

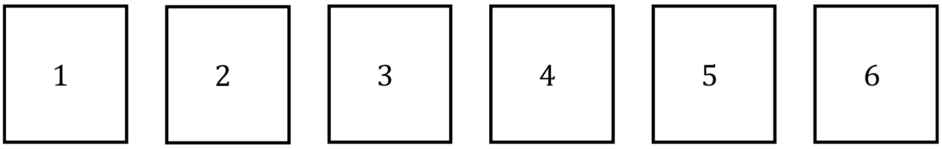

Figure 1.1 The outcome space of a fair die

This is the outcome space, the space of all possible outcomes, all possible results of throwing the die. Think of it as a set of paving stones numbered 1 to 6. It allows us to calculate the probability of events. We are going to use it to evolve rules for manipulating and calculating probabilities. The rules are useful for more complicated problems where working with the outcome space is cumbersome. However, even then, if you are uncertain about a solution, the problem can often be elucidated by studying the outcome space of a simpler, model problem. The outcome space is often referred to as the sample space or state space. ‘Outcome’ sounds more appropriate to me, but it is only a choice of words. The outcome space is usually written as the set $\{1, 2, 3, 4, 5, 6\}$ or more compactly as $\{1 : 6\}$. There is no significance in using numbers to label the outcomes, the faces of the die: they could as well be letters, $\{A : F\}$. For the present, we will persist with our 6 numbered paving stones.

Rule 0: The outcome space of a trial contains all the possible results of the trial.

We have numbered the rule as zero because it is fundamental, so fundamental that it does not mention probability!

When we throw our fair die, there are 6 equally likely outcomes. The chance that a particular face lands up is one in six or 1/6. The probability of, say, a 3 is 1/6. Does that give a sense of what we mean by probability? We shall define probability as the chance of some event happening.

The chance that one throw of the die gives a result somewhere in the range 1 to 6 inclusive is once in one throw or 1/1, which is 1. Though we do not know which it will be, we are absolutely certain that one of the faces will land up. The probability of any number other than 1 to 6 landing up is zero. We cannot throw a 7 because our die does not have a face with 7 on it and so it is not in our set of possible outcomes. The two ends of the probability range represent certainty: zero is certainty that a result will not happen and unity that it will. Everywhere in between, we are uncertain, but the likelihood of the result increases as we go from zero towards one.

We can write these things in a more compact form as

$$\Pr(3 \mid throw\ of\ fair\ die, W) = 1/6$$

$$\Pr(1:6 \mid throw\ of\ fair\ die, W) = 1$$

$$\Pr(not\ 1:6 \mid throw\ of\ fair\ die, W) = 0$$

Here, Pr means probability of what follows inside the brackets. The item before the vertical bar is the outcome, event, result or statement. The item after the vertical bar is a list of conditions that control the outcome; in this case, the conditions are one throw of a fair die and everything we know about the world, W, in which it is thrown. We read the vertical bar as the word 'given.' It is tempting to omit the conditions after the bar for brevity or laziness, but we do so at the risk of mistaking the conditions and reaching erroneous results. The event and the conditions together, that is everything together inside the brackets, are sometimes called the proposition.

Rule 1: Probability is conditional on all that we know about the event and its environment.

The probability is distributed over the outcomes. If each outcome is represented by a paving stone in Figure 1.1, we can signify its probability as the volume of water in a bottle standing on it. We have 6 paving stones with 6 identical bottles. Each bottle is $1/6^{th}$ full of water. Take, for example, the bottle on outcome 5. It is $1/6^{th}$ full; if it were completely full, the probability would be 1 and the only possible outcome of throwing the die would be 5. We could make this true by putting a 5 on every face of the die. Changing the numbers on the faces changes the probabilities of the outcomes. Thus, erasing a dot of the 5 to make it into a 4 will empty the water from the bottle on outcome 5 into that on outcome 4 and the probability of throwing a 4 becomes 2 × 1/6 = 1/3. The probability of any other outcome remains 1/6. Alternatively, we could decide that if we throw a 5, we throw again until we throw a number that is not 5. That is the same as taking the water from bottle 5 and dividing it equally amongst all the other bottles. The probability of 5 goes to zero and the probability of each of the other numbers increases to 1/5.

Rule 2: Probability is a conserved quantity. When we change the outcomes, remove them or combine them, we re-distribute the probability. The total always adds to one.

Apart from demonstrating rule 2, the examples in the last paragraph have different conditions on the probability; they would appear in the place of '..' in $\Pr(6 \mid .., W)$. If you write those statements in full, you will find the condition makes the expression cumbersome. However, you should also see that it is important as it also changes the values of the probabilities!

If we threw the die, a fair one with six faces labelled 1 : 6, thousands of times, we would expect each face to land up in about 1/6th of all the throws. If we can, in reality, in some kind of numerical simulation or in imagination, repeat the experiment many times, we expect the fraction of the total that comes up 6 to approximate the probability of throwing a 6. Probability relates to, but is not equal to, the real (or imagined) frequency of the particular result.

Rule 3: If we repeat an experiment many times, the fraction of the total number of results approaches the probability of that result as the total number of trials becomes very large.

Rule 3 can be useful for understanding probability and estimating a probability, especially from observations.

3 Addition of outcomes and addition of probabilities

It should be clear from the above that the probability of an event that is a combination of outcomes is the sum of the probabilities of the component outcomes. For example, the probability of throwing either 1 or six is the sum of the water in the bottles on the paving stones numbered 1 and 6. Thus,

$$\Pr(1\ or\ 6 \mid throw\ of\ fair\ die, W) = \Pr(1 \mid throw\ of\ fair\ die, W) + \Pr(6 \mid throw\ of\ fair\ die, W) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

The 'condition' as we have written it is becoming cumbersome. From now on we will abbreviate '$throw\ of\ fair\ die, W$' to '$1f$'. Remember what it means!

The probability of throwing a 3 is 1/6. The probability of throwing anything other than 3, we might call it *not* 3 is the sum of the probabilities of 1, 2, 4, 5 *or* 6; that is 5/6. More generally, if we think of some outcome X, the sum of the probabilities of outcome $X\ or\ outcome\ not\ X$ is unity.

Rule 4: The sum of the probability of an outcome X and the probability of outcome $not\ X$ given the same conditions is unity.

Consider some more complicated cases, ones that include more than one of our paving stones. The probability that our throw comes up 1, 2 or 3 is the sum of the water in the bottles on stones 1 to 3.

$$\Pr(1 : 3 \mid 1f) = \Pr(1 \mid 1f) + \Pr(2 \mid 1f) + \Pr(3 \mid 1f) = 1/2$$

The probability of throwing an even number is

$$\Pr(even \mid 1f) = \Pr(2 \mid 1f) + \Pr(4 \mid 1f) + \Pr(6 \mid 1f) = 1/2$$

The probability of ({1 : 3} or *even*) is not the sum of the last two probabilities; that would include the probability of 2 twice. We could obtain the answer from our outcome space by summing the water in the bottles on stones 1, 2 and 3 and the water in 4 and 6. We can take the water from bottle 2 only once! Thus,

$$\Pr(\{1 : 3\}\ or\ \{even\} \mid 1f) = \Pr(1 : 3 \mid 1f) + \Pr(even \mid 1f) - \Pr(\{1 : 3\}\ and\ (even) \mid 1f)$$

The last term is the probability of outcomes included in {1 : 3} and in $\{even\}$. We might think of it as the overlap between the two sets of outcomes or their joint occurrence.

Now, we can formulate a general rule: if A and B are two results of some experiment given conditions X, then

$$\Pr(A \; or \; B \mid X) = \Pr(A \mid X) + \Pr(B \mid X) - \Pr(A \; and \; B \mid X)$$

Rule 5: Probabilities of independent events add. When they are not independent the probability of the two events occurring together, the 'joint probability,' is subtracted.

Joint probability, the probability of two results coming together, is very important, something that we often have to deal with and, when we do, the dependence or independence of the outcomes is crucial.

4 Joint probability

We wrote the joint probability of the two events, A and B, as $\Pr(A \; and \; B)$. However, since joint probability arises so often, we abbreviate the nomenclature to $\Pr(A, B)$. We start with an example where the events, A and B, are independent. If we throw two dice together, a red one and a blue one, what is the probability that the red one lands with face 2 up and the blue on with face 3 up? The outcome space of for the two dice is shown below.

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

Figure 1.2 The outcome space for two fair dice

The first number on the paving stone is the result for the red die and the second that for the blue. Thus, for the whole first row, the red die result is 1; the second row is, the red gives 2 and so on. The columns correspond to results from the blue die: first column is 1, second is 2, … . There are now 6×6 = 36 paving stones, each with its bottle of water now filled to a level of 1/36th. It is clear from the outcome space that the answer is 1/36, $\Pr(2, 3) = 1/36$. However, the outcome space is already complicated for two dice; we need a general rule. The probability of the red die giving 2 is 1/6, that is the probability of the entire second row of the outcome space. Given that, the probability of the blue die comes up 3 is 1/6. We might write that as

$$\Pr(2,3 \mid 1f(R), 1f(B)) = \Pr(R = 2 \mid 1f) * \Pr(B = 3 \mid R = 2, 1f) = 1/6 * 1/6 = 1/36$$

Of course, in the example above, the result of the blue die is completely independent of that of the red one. We return to an earlier example with one die to explore the effect of dependence.

In the previous section, we looked at the probability of throwing a result $\{1 : 3\}$ **or** $even$, $\Pr(\{1 : 3\}\ or\ (even) \mid 1f)$, given one fair throw of a single die. Now consider the probability of $\{1 : 3\}$ **and** $even$, $\Pr(\{1 : 3\}, \{even\} \mid 1f)$, given one fair throw of a single die. Since there is only one even number in the set $\{1 : 3\}$ and that number is 2, the answer is obvioulsy1/6. We want to find a rule for obtaining that result. The probability of the throwing a value in the range 1 to 3 inclusive is

$$\Pr(1 : 3 \mid 1f) = \Pr(1 \mid 1f) + \Pr(2 \mid 1f) + \Pr(3 \mid 1f) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

The probability of that the throw is an even number conditional on the number falling in the range 1 to 3 inclusive is 1/3,

$$\Pr(even \mid \{1 : 3\}, 1f) = \Pr(2 \mid \{1 : 3\}, 1f) = \frac{1}{3}$$

Note that the answer is 1/3 and not 1/6. That is because we have excluded the possibility of the faces 4 to 6. All the water in the bottles on paving stones 4 to 6 has been transferred to the bottles on 1 to 3.

The probability of $\{1 : 3\}$ **and** $even$ is the product of the last two results, $1/3^{rd}$ of a half is $1/6^{th}$.

$$\Pr(\{1 : 3\}, \{even\} \mid 1f) = \Pr(1 : 3 \mid 1f) * \Pr(even \mid \{1 : 3\}, 1f) = \Pr(even \mid \{1 : 3\}, 1f) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$

We could have made the calculation in the reverse order. Calculate the probability that the throw is even and the probability of it falling in the range 1 to 3 conditional on it being even. The answer would be the same. Our general rule is

$$\Pr(A, B) = \Pr(A \mid B) * \Pr(B) = \Pr(B \mid A) * \Pr(B)$$

We have omitted the extra condition X after the vertical bar to emphasise the symmetry of the equation, but it must not be forgotten.

Rule 6: The probability of two things occurring, the joint probability of two results A and B, equals the probability of the one, A, times the probability of the other, B, conditional on A occurring.