BUSINESS STATISTICS

Using EXCEL & SPSS

Nick Lee & Mike Peters

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Kirsty Smy
Editorial assistant: Molly Farrell
Production editor: Nicola Marshall
Copyeditor: Neville Hankins
Proofreader: Andy Baxter
Indexer: Gary Kirby
Marketing manager: Alison Borg
Cover design: Francis Kenney
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound in Great Britain by Ashford Colour Press Ltd

MIX
Paper from responsible sources
FSC® C011748
FSC
www.fsc.org

First published 2016

Library of Congress Control Number: 2015939926

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-84860-219-9
ISBN 978-1-84860-220-5 (pbk)

5

INTRODUCTION TO PROBABILITY

CONTENTS

learning objectives

Here are the learning objectives for this chapter:

- ☑ Understand the probability of an event as the likelihood of the occurrence of that event, expressed by a number between 0 (no chance) and 1 (certain).
- ☑ Understand the difference between theoretical and empirical probability.
- ☑ Learn what a probability experiment, a sample space, an outcome and an event are.
- ☑ See how a tree diagram can represent multiple step experiments and how the counting rule can be used to determine the total number of outcomes.
- ☑ Understand how the counting rules regarding permutations and combinations can be used to determine how many permutations and combinations are possible from any given set.
- ☑ Learn how to assign a probability to an outcome using classical, relative frequency and subjective methods.
- ☑ Learn the relationship between the probability of an event and its complement, and the addition law.
- ☑ Begin to understand conditional probability, joint probability, marginal probability and the multiplication law.
- ☑ Learn about basic Bayesian concepts and Bayes' Theorem.

Probability is a concept that often baffles people; certainly I remember struggling with it as a student, and even as a lecturer when I first started teaching statistics. But it really does not have to be a mystery. Think about your day-to-day life: you are dealing with probability on an almost constant basis whether you know it or not. Cognitive psychology aside, every time you cross the road you automatically judge the likely probability of making it successfully. In fact, every time you make a choice of whether or not to act in a situation, you have somehow judged the probability of the outcomes. You might call it 'instinct', but it is really a rough probability judgement. While many scientists have explored failures in our ability to judge, in a general sense we humans are pretty good at judging the likely consequences of our actions – well, at least compared with other animals – even without prior experience of the action. Of course, our probability judgements may not be accurate, but at least we are making them. In fact, some say the most successful people are those who judge inaccurately, being over-confident of success and under-confident of failure. But I digress. The point I am making is that you automatically do on a daily basis many of the things I am about to introduce. All we need to do is put numbers on them.

In the business world, of course, we are often interested in putting numbers on probabilities. This helps decision makers with their decisions. Perhaps more importantly it helps them justify to their superiors exactly *why* they made a certain decision – especially if it is wrong. In the past, management has often been characterized as being largely 'gut feeling' – or instinctual probability judgements. However, in today's environment the ability to understand and work with probability figures is far more valuable in the main, allowing us to make good decisions clearly and consistently. Probability is also indispensable to statistics, as you will see in a few chapters. Basically, we cannot do any statistical testing without relying on an underlying base of probability,[1] because we are trying to *estimate* the chances of certain things happening in the real world, but using only a sample of data, like sampling a wine.

BASIC CONCEPTS IN PROBABILITY THEORY

Probability values are always indicated by a number between 0 and 1, with a value near 0 indicating a small likelihood and near 1 a large likelihood. Sometimes in other settings you can see this number represented in other ways, like a percentage or a ratio (e.g. 2:1, as in betting odds), but these are not formal representations of probability, and they can always be easily converted into a value between 0 and 1.

The simple rule to remember is that *as the number gets closer to 1, the event is more likely to happen*, and at 1 it is certain to happen.

There are two ways in which we can estimate this probability number. We can use logic to make a judgement if we want. For example, if you are into complicated fantasy games like some of my friends, you might have great experience with many-sided dice. You can estimate the probability of obtaining a particular value here easily. If the die is fair, the probability is always 1/the number of sides of the die. So for six sides the probability of getting a '5' is 0.167 (1/6), and for a 12-sided die (called a D12 by those 'in the know'), the probability of getting a 9 is 0.0833. This is called theoretical probability, and is what we usually roughly estimate automatically, without even knowing it. However, more often in a quantitative analysis situation we work with empirical probability, which is where we estimate the probability of an event based on past observations.

A probability experiment is a process which generates some kind of outcome, which we can use to estimate empirical probability. Rolling a die is a probability experiment, with the outcomes being each possible number on the die, but almost anything can be defined in this way, as long as it has discrete, definable outcomes (i.e. only one of the outcomes can occur at any one time). The sample space is the set of all possible outcomes for the experiment. Thinking about this as a *set* will be of considerable use to you as you move on through the book, so Back to Basics Box 5.1 introduces very basic set theory and terminology.

[1] Technically, I'm being over-simplistic here, because I'm ignoring the entire field of Bayesian statistics. However, this is far beyond the scope of this book in the main, and I encourage you later in your studies to find out more about this fascinating approach.

back to basics

Box 5.1 Fundamentals of Set Theory

Imagine we are playing a game of 'truth or dare' around a table at a party. There are six of us, and we each have a number. We can roll a six-sided die for a more random selection method than spinning an empty bottle. When your number comes up, it is truth or dare time!

Now, think about all the possible outcomes of the roll of a six-sided die. This is pretty obvious: they are 1, 2, 3, 4, 5, 6. Are you with me? Set theory begins with the relationship between a single *object* and a single *set*. If an object is a member of a set, we say that it is an element of the set. Consider the possible outcomes of the die roll as objects and the collection of all possible outcomes of the die roll as a set. The individual outcomes are thus elements of the set of all possible outcomes.

If an object o is an element of a set A we can use specific terminology to show this:

$$o \in A$$

The basic relationship between two sets is the subset, or set inclusion. For example, if all the members of set B are also members of set A, then B is a subset of A. The way to show this is:

$$B \subset A$$

Moving back to the die example, if set A is the possible die roll outcomes – which we can show enclosed in curly brackets like this: {1, 2, 3, 4, 5, 6} – and B is the set of possible outcomes from a four-sided die roll {1, 2, 3, 4}, then:

$$B \subset A$$

There are lots of other common set operations which will be of use to you later in this chapter, and in the book.

The union of the sets A and B is the set which contains all members of at least one of either A or B. This is denoted:

$$A \cup B$$

For example, if $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{9, 10\}$, then:

$$A \cup B = \{1, 2, 3, 4, 5, 6, 9, 10\}$$

The intersection of sets A and B is the set which contains all members of both A and B. This is denoted:

$$A \cap B$$

For example, if $A = \{1, 2, 3, 4, 5, 6\}$, and $B = \{5, 6, 7\}$ then:

$$A \cap B = \{5, 6\}$$

The complement of set A relative to set U is denoted as A^c. It means all the members of set U which are not members of set A. We use a set called U not B in this case, because complements are most often used when U refers to what is called a universal set. Different types of set theories have different universal sets, but for probability theory the universal set is all the possible events that could happen – which you might know as the sample space! So the complement of set $A = \{1, 2, 7\}$ relative to a set U of the sample space of a six-sided die experiment $= \{1, 2, 3, 4, 5, 6\}$ is $\{7\}$.

There are many other operations on sets, which may be introduced later on, but those above are the basics. You might also remember set theory being expressed in terms of diagrams of overlapping circles – called Venn diagrams. Don't worry, these will appear later on!

think it over 5.1

The word 'set' sounds a bit intimidating. One straightforward way to think of a set is to go shopping. You have two bags to carry the shopping home. Since you are a very organized person, you put all the freezer stuff in one bag and all the alcoholic drinks in another. The freezer bag contains ice cream, chicken balti ready meal, beef curry ready meal and pizza. The alcohol bag contains four cans of lager, four cans of cider, four cans of bitter and a bottle of whisky. In mathematical notation, let F be the set of frozen items and A be the set of alcoholic items:

F = {ice cream, chicken balti ready meal, beef curry ready meal, pizza}

A = {4 cans of lager, 4 cans of cider, 4 cans of bitter, 1 bottle of whisky}

U = all possible items you could have purchased

How would you write mathematically that pizza belonged to set F?

What does $A \cap F$ mean and does it have any members?

What does $A \cup F$ mean and does it have any members?

think it over 5.2

Consider the following.

Suppose there is a town with just one male barber and that every man in the town keeps himself clean-shaven, some by shaving themselves, some by going to the barber. It seems reasonable to imagine that the barber obeys the following rule: He shaves all and only those men in town who do not shave themselves.

Under this scenario, we can ask the following question: Does the barber shave himself? Or in mathematical language: Which set does the barber belong to? (Men who shave themselves or men who are shaved by the barber.)

The different possible outcomes in the sample space are called sample points. In the sample space for the six-sided die roll experiment used already, there are six sample points. However, it is not always easy to count up all the possible experimental outcomes. In some cases we have rules that can help us, which will be discussed below.

Multiple Step Experiments and Tree Diagrams

Imagine the common practice of measuring defects in production. The simplest check is whether a product which is produced has a defect or not. So, every time a product was finished, and rolled off the production line, we could employ a quality assessor to check it for any defects. A product check would be an experiment with two possible outcomes – defect or no defect. If defect was denoted D and no defect was denoted N, we could describe the sample space like this:

$$S = \{D, N\}$$

To illustrate the idea of multiple step experiments, imagine if we were producing a total of two products, so our experiment would consist of two product checks (i.e. a two-step experiment). The sample space for this two-step experiment would be a little larger than the single-step example above:

$$S = \{(D, D), (D, N), (N, D), (N, N)\}$$

To make this clearer, there are four elements here – each enclosed in normal brackets – and the elements are all enclosed in curly brackets. The first element of (D, D) indicates that there was a defect on the first check and a defect on the second. So in this case, it was easy to list all of the possible outcomes. But in many more complex situations it is not so simple. Fortunately, the *counting rule for multiple step experiments* allows us to figure out the number of experimental outcomes without listing them all out as above.

box 5.1

The Counting Rule for Multi-step Experiments

If we can describe any multi-step experiment as a sequence of k steps with n_1 possible outcomes for the first step, n_2 possible outcomes on the second step, up to n_k possible outcomes at the kth step, the total number of experimental outcomes is:

$$n_1 \times n_2 \times ..... \times n_k$$

Thus, even if we could not have listed the two-step quality-check experiment above, it is easy to work out the possible outcomes by considering the first quality check ($n_1 = 2$), followed by the second check ($n_2 = 2$). The counting rule tells us that the total number of outcomes is $2 \times 2 = 4$. If we were to check three products, the total number of outcomes would be $2 \times 2 \times 2 = 8$. And so on.

Another way of doing this, which is often used when we want to look at the differences between the individual outcomes as well as count the possible number of them, is to use a tree diagram. This diagram is a way of graphically representing a multiple step experiment, and it shows each possible combination of outcomes, which can help us understand the range of possible outcomes from a multiple step experiment.

As an example, let's consider the following situation, which will be used in various places throughout this chapter. Imagine you have an assignment to plan for your course. You can split it into two stages: research and writing. The stages are sequential in this instance, because you want to complete all your research before beginning to write. However, you cannot predict the exact time each of those stages will take. Even so, your prior experience in the past suggests that you can take anywhere from 10 to 21 days to do your research, and anywhere from 5 to 14 days to write it all up. You have a deadline in 4 weeks (28 days), and you need to be done by then. For simplicity's sake at this point, you can set this up as a two-stage experiment, each with three possible outcomes corresponding to the shortest and longest times, as well as a time in the middle. Thus, the first stage (research) could take 10, 16 or 21 days, and the second (writing) could take 5, 10 or 14 days.

Applying the counting rule above is simple and shows there is a total of $3 \times 3 = 9$ possible outcomes, and we can describe them numerically as well. For example, the outcome (10, 5) would indicate that the research stage took 10 days and the writing took 5, which would be a total of 15 days for the project – well within the deadline! Figure 5.1 shows a tree diagram which sets out all the possible outcomes.

The tree diagram allows you to create the sample space quite easily – from the sample point column. From the diagram you can see that the completion time for the assignment ranges from 15 to 35 days, and six of the nine possible outcomes result in a completion time under 28 days. Of course, we have no

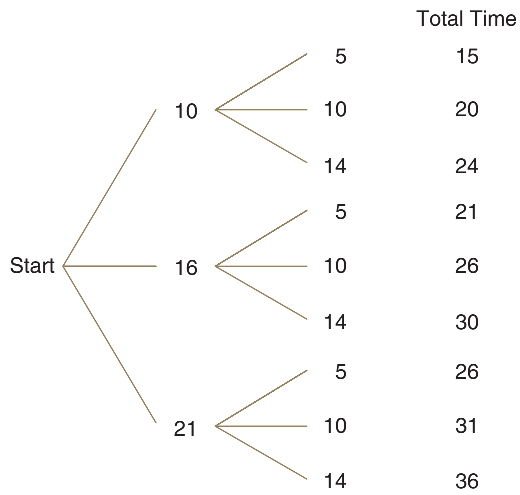

Figure 5.1
Tree Diagram for the Assignment

idea at this point of the actual probability of any of these particular outcomes, so right now we do not know if they are equally likely to occur. As a result, we would need more information before assessing the likelihood of the assignment being completed in under 28 days.[2]

Permutations and Combinations

When I was a first-year undergrad, one of my proper genius friends (hi Todd!) was already doing a second-year maths paper called 'permutations and combinations'. I thought it sounded really difficult, but, really, it is all about counting the possible number of outcomes or objects – which you might think is pretty useful right now. We have already worked out a simple counting rule, but what do we do in more complex situations? Imagine you were an auditor checking company records for accuracy. As already mentioned in Chapter 1, most of the time auditors check a smaller sample of records, not the whole lot. So if you had a set of six bi-monthly expense reports from salespeople, and you could check only three of them, how many possible different combinations of three records could be selected from the set of six? To work this out we need the counting rule for combinations.

box 5.2

The Counting Rule for Combinations

The number of possible combinations of k taken from a set of N objects is given by:

$$C(N, k) = {}_{N}C_{k} = {}^{N}C_{k} = \binom{N}{k} = \frac{N!}{k!(N-k)!}$$

OK, I know that looks complex, but it really just shows different ways it can be notated. The actual formula is on the right hand side, the fraction with the exclamation marks.

One important term you might not have come across is the factorial, which is indicated by the exclamation mark ! in the formula. Factorials are really quite simple, and just mean that for any non-negative

[2] In fact, rather than spending time working this out, you'd probably be better off just getting on with doing the assignment!

integer n (i.e. any whole number from zero to infinity inclusive) the factorial is the product of all the positive integers less than or equal to n. What does this mean? Well…:

$$4! = 4 \times 3 \times 2 \times 1 = 24$$

By the way, you might be wondering what 0! is? Well, it is one. Look, it just is, OK? Oh, go on then – if you want an explanation, check Above and Beyond Box 5.1 (I haven't done one of these for a while).

above and beyond

Box 5.1 Why 0! = 1

Factorial 0! = 1 because it represents zero multiplied by nothing; the result of multiplying no numbers is called the empty product and has a numerical value of 1. Why is this? Well, there are lots of different definitions and justifications of why this is the case, but I am going to show you only one of them. Imagine the multiplication of 3 × 5 × 5, which can be represented by:

$$\text{prod}((3 \times 5 \times 5)) = \text{prod}((3 \times 5)) \times 5 = \text{prod}((3)) \times 5 \times 5 = \text{prod}(()) \times 3 \times 5 \times 5$$

Now of course, the only logical thing which can be in those brackets is a one, making:

$$1 \times 3 \times 5 \times 5$$

If a zero was in there, the equation would all of a sudden fall apart and equal zero.

It kind of makes sense if you think about it.

Moving on to our combinations problem, remember that we have six expense reports, and can select three. Thus, $N = 6$ and $k = 3$. So inputting these into the counting rule for combinations results in:

$$C(6,3) = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1) \times (3 \times 2 \times 1)} = \frac{720}{36} = 20$$

So there are 20 possible outcomes of the experiment of selecting three reports from a group of six. You will find that the numbers quickly get extremely large. For example, if you think about a lottery – say the first one introduced in New Zealand when I was a kid, which had 6 balls being selected from 40 – the possible combinations of 6 balls from 40 would equal 3838380 – cunningly just more than the population of New Zealand at the time of its introduction (not an accident I imagine). So you had nearly a 1 in 4 million chance of winning the top prize at that time. Add a few more balls to the total, increase it to 45 for example, and the possible combinations leap to over 8 million, and at 50 balls the number is nearly 16 million!

think it over 5.3

Which is the best strategy for improving the probability of winning the lottery?

Strategy 1 – stick with the same numbers.

Strategy 2 – use different numbers each time.

If the probability of winning the UK National Lottery is approximately 1 in 14 million and the population is roughly 60 million, why are there not four winners per draw?

Now, imagine you are interested not just in a particular combination, but also in the *order* that the combination was selected in. So the same set of k objects is a different outcome when it is selected in a different order (unlike the combinations situation above).

box 5.3

The Counting Rule for Permutations

The number of possible permutations of k taken from a set of N objects is given by:

$$P(N, k) = {}^{N}P_{k} = k!\binom{N}{k} = \frac{N!}{(N-k)!}$$

You should be able to see that the denominator of the permutations formula (the bottom part of the fraction) is less than that for the combinations formula, therefore the number of permutations for a given number of objects will be greater. This makes sense, since every selection of k objects can have k! different orders.

Let's stick with the auditing example, where you need to select three expense reports from six to check. If you are concerned with the order of selection you will want to use the permutations rule, not the combinations one:

$$P(N,K) = \frac{6!}{(6-)3!} = \frac{6!}{3!} = 120$$

So 120 outcomes are possible for the experiment of randomly selecting three expense reports from a group of six, when the order matters. As you can see, the numbers here will quickly get huge. In the New Zealand lottery example, if you have to get the correct exact order of the 6 numbers from the 40, your chances of winning are now 1 in 2763633600. Yes, that is nearly 1 in 3 billion.

Assigning Probabilities to Outcomes

Up until now, we have been working purely within a somewhat abstract theoretical framework, so let's come back to the real world. A key question we would probably be interested in is how we assign probabilities to experimental outcomes in the real world. This is the problem we had with the assignment example earlier – the tree diagram was interesting, but almost useless as a decision tool without some probabilities assigned to the outcomes. Even so, before we can even start assigning probabilities, we need to consider two fundamental conditions:

- The probability of *any given outcome* must be between 0 and 1 inclusive. This makes sense, because if you remember I said earlier that 0 can represent no chance of an outcome and 1 could represent certainty. So every possible probability must range between these extremes (no possible outcome can have less chance than none, or more chance than certainty, surely?!).
- The sum of the probabilities of *all possible outcomes* must equal 1. This makes sense, because if the set contains all possible outcomes, then there must be a certain chance of one of them happening, no matter how small the chance of any individual one is.

There are three main approaches used to assign probabilities, and each is appropriate in different situations. The simplest method is the classical method, which is the best one to use if you are sure that all of the outcomes are equally likely. In this case, if n represents the number of outcomes that are possible, then the probability of each individual outcome is simple 1/*n*. The bonus here is that the two conditions above are automatically fulfilled. A very simple example is the die roll already discussed. There are six possible outcomes for a roll of a six-sided die, and if the die is fair each outcome is equally likely. The probability of any outcome is therefore 1/6, or 0.167.

The classical method used above was a theoretical method. In other words, I did not need to roll the die to estimate the probabilities, I had a theory which told me that there were six outcomes, each of which was equally likely. In that case, this was justifiable, but of course it is not often that we are certain about such things. In many cases, we have data from prior experience or research which suggests the possible probabilities. In such cases you can use the relative frequency method. This is an *empirical* method, since it relies on data from the real world to estimate probability.

Table 5.1 Data for errors in first-year test

Student name	Number of errors in paper
Dagmar Baggins	8
Carole Reddy	2
John Cardigan	8
Rowena Yates	4
Jon Fall	3
Anne Surgeon	0
Max Power	7
Brylie Russ	3
Laura MacKenzie	0
Annamarie Green	6
Geert van Damme	1
Carl Old	4
Jack Junior	0
Lucy Textor	1
Caroline Rider	2
Keesje Stars	7
Huw Brewer	3
Charlie Danger	6
John Radd	2
Trecia Painter	5

Take the data in Table 5.1, for example, which shows the data from a quick study of error rates in the first quantitative analysis test of the year. You can see that out of the sample of 20 tests that I checked, there were no errors in three of them, ranging to eight errors in two of them. 'Not bad' you might think. But then I might tell you that there were only 10 questions. Anyway, we can use this data with the relative frequency method to calculate probabilities very easily. For example, the chance of getting no errors can be assigned a probability of 3/20, or 0.15, and that of getting eight errors is 2/20 or 0.1. Again, this is pretty simple (if you have the data) and automatically satisfies the two conditions above.

think it over 5.4

I teach a similar group with the same number of students. Can I use the above data to predict precisely the number of students who will make no errors in this new group? Give a reason for your answer.

Things get more complex if we have little good data, or if we cannot assume equal probabilities for outcomes. Unfortunately, this is quite common in reality, and means the previous two methods are of no use to us. In such cases we must assign probabilities by using the subjective method. If this sounds a little to you like 'make it up' then you would be right. The subjective method is basically the use of any available information, including prior experience, others' experiences and even your own assumptions, opinions or intuitions. Of course, everyone's subjective estimates might be different, so two different people might decide on different probabilities for the same event. In fact, this is what the gambling industry depends on – I might judge the odds of my favourite football team winning more generously than the bookmaker. If I am correct then I make money, but bookmakers depend on the assumption that they will be more correct more often, over the course of many bets by many people, and thus they will come out ahead.

With this in mind, consider the following example. I am a big cricket fan, and I enjoy the regular contests for the 'Ashes' between England and Australia. Now, being born in Wales and living in New Zealand for a long time, I really do not want Australia to win, but I know plenty of Australians who do rather enjoy an Australian victory and are not shy of letting the world know about it. Imagine if, in 2009, I bet one of them £10 that England would win the Ashes. Two outcomes are possible – England win and regain the Ashes so traumatically lost in 2006–7 (designated O_1), or Australia win and retain the Ashes (designated O_2). Before the beginning of the series, I believed the chances of an England win to be 0.6, and thus I set $P(O_1) = 0.6$ and $P(O_2) = 0.4$. On the other hand, my betting buddy, being an Aussie, might set $P(O_1) = 0.1$ and $P(O_2) = 0.9$. Thus, he is very optimistic about Australia's chances. This is clearly a situation where at least one of us is letting their allegiances cloud their judgement – or perhaps both of us are?

Either way, it can be seen that in this case we have both assigned probabilities that adhere to the two fundamental conditions for assigning probabilities. This is important to keep in mind, because even when using subjective assignment, it is logically impossible to set probabilities which do not have a value between 0 and 1 inclusive, and where the sum of all outcomes does not equal 1.

As a final note, the methods above are not mutually exclusive. Instead, one can combine empirical data with subjective experiences and intuitions to estimate probabilities, or combine classical estimates with subjective probabilities. Doing so may make use of all the possible information, and result in more accurate estimates.

THE PROBABILITY OF EVENTS

You probably think you know what an 'event' is, right? *Wrong!* For the purposes of probability theory, an event has a specific definition. First, remember that a sample point is an outcome of a probability experiment. An event is simply a collection of sample points. Why talk about events? Events are useful because in many cases we are more interested in the probability of a set of sample points than any single one, as you will soon realize. For example, if you have an assignment due in 28 days, you are more interested in the probability of the set of sample points representing 28 days or less for assignment completion than any particular amount of days (e.g. 27, 26, 15 and so on). If you refer back to the example in Figure 5.1 you can see that six of the nine sample points give us a hand-in time less than

28 days away. These points are: (10, 5), (10, 10), (10, 14), (16, 5), (16, 10), and (21, 5). So, we use C to indicate the event that the assignment is completed in less than 28 days, shown by:

$$C = \{(10, 5), (10, 10), (10, 14), (16, 5), (16, 10), (21, 5)\}$$

Thus event C occurs if any single one of those sample points occurs. It is pretty easy to calculate the probability of an event. It is simply *the sum of the probabilities of the individual sample points in the event*. Of course, this means you have to define probabilities first for each sample point (which has not been done for this example). How you do this is a key decision, and the different methods were given above, but once you have done this, it is quite simple to define the event and sum the individual probabilities of the sample points.

Or perhaps I spoke too soon. Above it was easy, but what if you have hundreds of sample points? In principle it would still be simple, but in practice this would be a bit of a pain. So, we have a set of really helpful principles which can be used to work out the probability of an event, even if we do not have all the individual sample point probabilities. And it is at this point that your earlier digression into set theory will come in handy – but this time there will be pictures!

The Complement of an Event

The simplest situation to start with is the complement of an event A, which is defined as all of the sample points which are not in A. The complement of A is denoted by $\bar{A}$, and I can most easily show it with a Venn diagram, which is another way of looking at sets. You might have first come across Venn diagrams at school; they are made up essentially of circles and boxes with shading, to represent sets. Figure 5.2 shows the complement nicely. The shaded box is the entire sample space for the experiment, and the white circle represents event A. Thus, the complement $\bar{A}$ is all of the sample points not in A, meaning it is the complement of A.

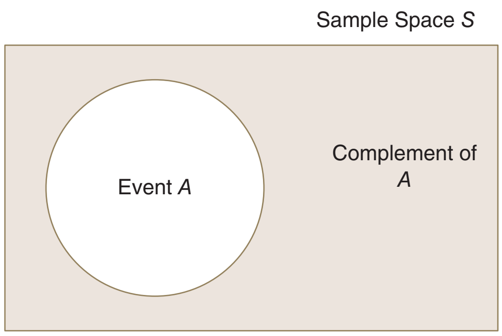

Figure 5.2 The Complement of *A* (shaded)

If you think about it, it is always true that in any probability situation, either event A or its complement must occur, and therefore:

$$P(A) + P(\bar{A}) = 1$$

So, if you solve this equation for $P(A)$ you get:

$$P(A) = 1 - P(\bar{A})$$

Thus, getting the probability of any event is simple, if you know the probability of the complement of an event. In fact, if you know one, you must know the other. But when would this situation actually occur in reality? It is not as uncommon as you might think. Imagine you are trying to compute the probability of a customer making a purchase once he or she walks into your store. Your records show that 60% of those who enter the store make a purchase. You can use A to represent the event of a sale and $\bar{A}$ to represent the no-sale event. In the absence of any other information, you must set the probability of a sale event at 0.6, thus $P(A) = 0.6$. Using the equation above you can easily see that the event of no sale has a probability of 1 – 0.6, and therefore $P(\bar{A}) = 0.4$.

The Addition Law

Before moving to the addition law, I want to recap some more key set theory concepts which are relevant, and transfer them to Venn diagrams. First, remember that an event is simply a special type of set, so for now I am going to stop talking about sets and use the word event, and, instead of elements, I am going to use the term sample point. If you get confused with how this relates to the content of Back to Basics Box 5.1, just swap the terms around for a while. Now, remember that the *union* of any two events is the event which contains all sample points of both events, so given two events A and B, the union of A and B is all sample points in A and B, and is given by $A \cup B$. In the Venn diagram in Figure 5.3 the union is shaded. You can see the circles overlap, and this means that there are some sample points in both A and B, but this is not a necessary condition for a union.

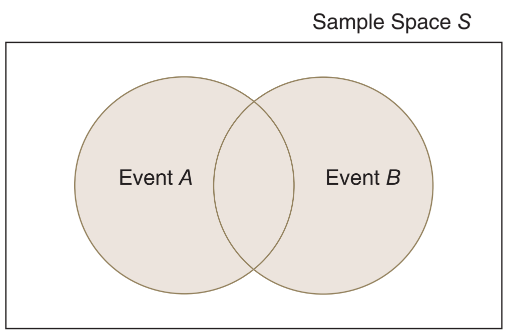

Figure 5.3 The Union of *A* and *B* (shaded)

Now, recall the intersection idea. The *intersection* of A and B is the event which contains only the sample points which belong to *both* A and B, given by $A \cap B$. Figure 5.4 shows a Venn diagram representing the intersection relationship, with the intersection shaded.

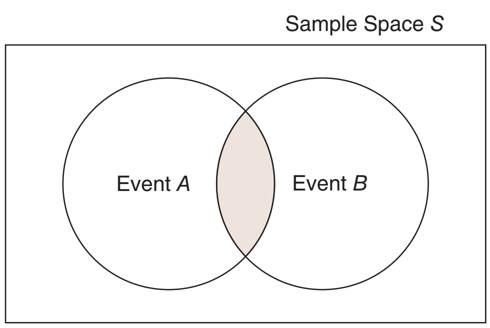

Figure 5.4 The Intersection of *A* and *B* (shaded)

To understand the addition law, you need to understand the concept of union and intersection. But first, the addition law itself, as follows.

box 5.4

The Addition Law

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

think it over 5.5

If events *A* and *B* did not have any sample points in common, how would you write the formula in Box 5.4? Draw a Venn diagram to illustrate the situation where there are no sample points in common. What is the statistical term used to define this situation?

This makes sense intuitively because we need to account for the sample points which are in both *A* and *B*. If we just add *A* and *B* together we will double count all those sample points in both (if there are any). So to account for this we must subtract the intersection $A \cap B$ to ensure those points are only counted once.

Let's work through this in an example. Imagine you are a sales manager with 10 sales representatives out in the field. At the end of the month, four of your reps fell short of the target, three of them did not complete their reports properly, and two fell short of the target and also didn't complete their reports properly. Let:

A = the event of falling short of target

B = the event of not completing the reports properly

Looking at the relative frequencies we can determine the following probabilities:

$$P(A) = \frac{4}{10} = 0.3$$

$$P(B) = \frac{3}{10} = 0.4$$

$$P(A \cap B) = \frac{2}{10} = 0.2$$

Your management style is such that you consider that both missing your target and not filing your reports properly are worthy of a poor rating, so you are interested in the probability of $A \cup B$, which is the probability of getting a poor performance rating. To work out this probability we can use the addition law as shown above:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

You can easily substitute the probabilities above into this as follows:

$$P(A \cup B) = 0.3 + 0.4 - 0.2 = 0.5$$

This tells us that if we were to select a sales rep randomly from your team, there would be a 0.5% chance that the rep would have received a poor rating.

Conditional Probability

I often tell stories of when I first became a lecturer and had to teach introductory quantitative analysis. Writing this chapter reminds me of the exact time when I realized that I had a lot of work to do to get up to speed: it was when I picked up the notes for the lecture on conditional probability. I remember thinking 'this is really tough', and I am hoping that I can explain it in such a way that you will find it much easier than I did.

Conditional probability is based on the idea that, in some cases, the probability of an event is based on whether some other related event has already occurred. In this case, if we are interested in an event A which has a probability $P(A)$, and if we know that some related event B has occurred, then we can use this information to calculate a new probability for A, called a conditional probability, and shown as $P(A|B)$. The vertical line | means that we are thinking about the probability of event A given that B has already occurred.

I would like to work this through with an interesting example drawn from my research. One of the things I am interested in is the decisions managers make about whether or not they should punish unethical employees. There are lots of interesting things which could influence this decision, but to avoid any heated arguments about what they might be, I am going to look only at the past performance history of the employee in this example. It is often thought that managers are reluctant to punish high-performing employees for unethical behaviour, whereas they are much quicker to punish low performers. Conditional probabilities could be used to analyse this situation quite nicely. Imagine we had data for 1000 unethical employees. Let:

H = the event that the employee is a high performer

L = the event that the employee is a low performer

P = the event that the employee is punished for an unethical action

$\bar{P}$ = the event that the employee is not punished for an unethical action

Table 5.2 Employee punishment data

	High performer	Low performer	Total
Punished	530	210	740
Not punished	190	70	260
Total	720	280	1000

Using this data, you can define the following probabilities using the skills you have already learnt. If you get confused, look carefully at the symbols and go back to the relevant section. First, the probability that a randomly selected employee is a high performer and is punished is the intersection of the high-performer and punish events, clearly seen in the relevant cell of Table 5.2, namely the number 530:

$$P(H \cap P) = \frac{530}{1000} = 0.53$$

Note especially that the non-punish event is actually the complement of the punish event, denoted of course as $\bar{P}$. So the probability that a randomly selected employee is a high performer and not punished is:

$$P(H \cap \bar{P}) = \frac{190}{1000} = 0.19$$

The probability of a randomly selected employee being a low performer and punished is:

$$P(L \cap P) = \frac{210}{1000} = 0.21$$

The probability of a randomly selected employee being a low performer and not being punished is:

$$P(L \cap \overline{P}) = \frac{70}{1000} = 0.07$$

Each one of these probabilities is an intersection of two specific events, and is called a joint probability. However, this information alone is not enough to work out the conditional probability. In fact, superficially it looks likely that being a high performer makes you more likely to be punished. But, similar to Simpson's paradox in Chapter 4, this might be misleading because what we are really interested in is the conditional probability. Table 5.3 summarizes this information and provides crucial information to work out the conditional probability. The values in the main body are the joint probabilities. The values in the row and column marked 'total' are called marginal probabilities (because some consider the outside row and column to be the 'margins' of the table) and represent the probabilities of each separate event, and are found by summing up the joint probabilities in the relevant row or column.

Table 5.3 Joint probability table for employee punishment

	High performer	Low performer	Total
Punished	0.53	0.21	0.74
Not punished	0.19	0.07	0.26
Total	0.72	0.28	1.00

To calculate the conditional probability of an employee being punished given that he or she is a high performer, namely $P(P|H)$, we recall that we are dealing only with the probability of punishment for the high performers. There were 720 high performers, and 530 of them were punished for unethical behaviour. It is simple to work out the conditional probability in this situation: $P(P|H) = 530/720 = 0.74$. So, given that an employee is a high performer who commits an unethical action, there is a 74% chance of their being punished for it.

That was easy, because I had the frequency data in Table 5.2 to work with. But now I am going to show you how to work it out with only the event probabilities to work with. First I will show you the basic principle. If $P(P|H) = 530/720 = 0.74$, try dividing both the numerator and denominator of the fraction by the total number of employees in the sample (1000):

$$P(P \mid H) = \frac{530}{720} = \frac{530/1000}{720/1000} = \frac{0.53}{0.72} = 0.74$$

So the conditional probability $P(P|H)$ can also be computed as 0.53/0.72, and if you look at the joint probability table (Table 5.3) you will see that 0.53 is the joint probability of P and H, namely $P(P \cap H)$, and 0.72 is the marginal probability that a random employee is a high performer, $P(H)$. So in other words the conditional probability $P(P|H)$ can also be computed as the ratio of the joint probability $P(P \cap H)$ to the marginal probability $P(H)$, or:

$$P(P \mid H) = \frac{P(P \cap H)}{P(H)} = \frac{0.53}{0.72} = 0.74$$

This leads us to a general formula for conditional probability, which I bet you could work out for yourself anyway.

box 5.5

Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

think it over 5.6

How could you show the conditional probability $P(A|B)$ with a Venn diagram?

So, moving back to the problem in hand, the marginal probability of being punished is 0.74, whether or not the employee is a high or low performer. The issue is whether or not you are more likely to be punished if you are a low or high performer. Thus the key is to compare the conditional probabilities $P(P|H)$ and $P(P|L)$. In other words, what is the probability of punishment given the employee is a high performer, and what is the probability of punishment given the employee is a low performer? If the two conditional probabilities are equal, there is no evidence for any bias of managers towards high performers, and a difference would indicate some bias.

Of course, you already worked out $P(P|H)$ as 0.74, so why not use the conditional probability formula to work out $P(P|L)$? Let's do so:

$$P(P \mid L) = \frac{P(P \cap L)}{P(L)} = \frac{0.21}{0.28} = 0.75$$

What conclusion should you draw? Well, even though the probability of punishment is higher if you are a low performer than a high performer, this difference is very small indeed. You would be on safer ground determining that there was probably no real difference. This result might raise one or two questions for you, for example how big does a difference have to be before we can conclude there is a difference? Indeed, this question has vexed the minds of many mathematical experts over the years, and is the foundation of *statistical inference*, which I will discuss later in the course of this book.

One final point to make about conditional probability is that not all events are dependent on others. Above, we were making the assumption that the probability of the punishment event (event P) is influenced by whether or not the employee was a high performer. If we find that the conditional probability of one event A given another B has occurred is different from the overall probability of A – expressed as $P(A \mid B) \neq P(A)$ – we say that A and B are dependent events. However, if this is not the case, and the probability of event A is not changed by the prior occurrence of event B, we say that A and B are independent events. In this case $P(A \mid B) = P(A)$.

The Multiplication Law

Previously, you saw that you could use the addition law to discover the probability of the union of two events. Here, I will show you how the multiplication law can be used to find the probability of the *intersection* of two events. The multiplication law is founded on conditional probabilities, which can be quite easily seen if you look at the conditional probability equation in Box 5.5. More specifically, if you multiply both sides of the equation in Box 5.5 by $P(B)$, you would get the following.

box 5.6

The Multiplication Law

$$P(A \cap B) = P(B)P(A \mid B)$$

which is the same as:

$$P(A \cap B) = P(A)P(B \mid A)$$

The multiplication law expresses that the probability of one event depends on the occurrence of the other. Thus the probability of both events occurring is dependent on the probability of one or the other occurring as well. So, in other words, the probability of A and B both occurring (i.e. $A \cap B$) is the probability of B multiplied by the probability of A given that B has occurred.

Strictly speaking, though, the multiplication law above only applies to dependent events. When events are independent we do not need to worry about conditional probabilities because $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$. Thus, you could easily substitute these into the equation above to get the multiplication law for independent events, which is $P(A \cap B) = P(A)P(B)$.

Some of you might be wondering why we need the multiplication law to find the intersection, since previously we used the data (as presented in Tables 5.2 and 5.3) to get the intersections. Well, in many cases we do not have that information, but instead we have probability information instead, so we can use the multiplication law.

BAYES' THEOREM AND BAYESIAN CONCEPTS

Well, that conditional probability was not so hard, was it? What I am going to introduce briefly now as the final section of this introduction to probability is Bayes' Theorem. It was about this point in my early lecturing career that I really got completely lost. It really took some serious efforts to begin to understand it. But, persevere I did, and hopefully now you'll reap the benefit of that![3]

Remember, one of the key concepts of conditional probability is that we can revise probabilities of a given event if we know about an earlier event happening. This concept is extended in the Bayesian approach. Oftentimes in the real world, we begin our probability analysis with a given estimate of probability – which we can call the prior probability estimate. Subsequently, it is common to receive additional information, like maybe a new report, or a piece of research we conduct or stumble upon. We can use this new information to *revise* our prior probabilities, to what are called posterior probabilities. We use Bayes' Theorem to calculate the posterior probabilities using the new information.

Using Bayes' Theorem can get quite complex and unwieldy, so I am going to go through it in a step-by-step manner, using an example. Imagine you are a human resources manager in a big international investment bank called Metrobank, with offices in the UK and the USA. In all, 42% of your staff work in London and 58% work in New York. Thus, if I was to select an employee at random, I would assign prior probabilities in the following manner: the probability of the employee being from London as $P(A_1) = 0.42$ and the probability of the employee being from New York as $P(A_2) = 0.58$. Now, imagine that the number

[3] Interestingly, concepts based on Bayes' Theorem – or what are now called 'Bayesian' approaches – have been influential in many ways since 1900 (including cracking the Enigma code in the Second World War), and there is now a real rise in Bayesian analysis in management and business research. It is well beyond the scope of this book to cover Bayesian analysis, but you might be interested in reading further on this as you move on in your career.

of employees who get their huge bonus for making the bank massive amounts of money differs across each location, with 12% of London employees making the bonus and 17% of New York employees making the bonus. If event *B* is the event that an employee makes a bonus, and *N* the event that an employee does not, we can work out the following conditional probabilities:

$$P(N \mid A_1) = 0.88 \qquad P(B \mid A_1) = 0.12$$

$$P(N \mid A_2) = 0.83 \qquad P(B \mid A_2) = 0.17$$

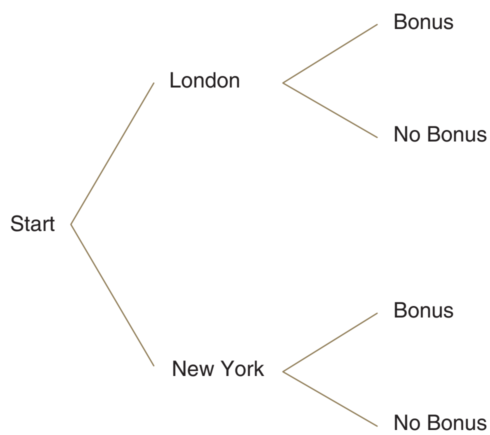

Figure 5.5 Tree Diagram for the Human Resource Example

Figure 5.5 shows a tree diagram for the process of your pulling an employee file at random, discovering whether it referred to a New York or London employee, and then seeing whether or not the employee made the bonus or not. This process can be represented as a two-step experiment with four outcomes (two bonus and two non-bonus). Each of the four outcomes is the intersection of two events, and this means the multiplication rule you learnt earlier should come in very handy, right? For example:

$$P(A_1, N) = P(A_1 \cap N) = P(A_1)P(N \mid A_1)$$

Figure 5.6 shows a probability tree incorporating this information and can help us work through the key issues. The probabilities shown at the first stage are the priors, and at the second stage the conditional probabilities are shown. Finding the probability of each individual outcome is simply a case of multiplying out the probabilities along the relevant branch.

Now, imagine you are at a bar having a drink (mine's a single malt if you're asking), and you are talking to a stockbroker who says she's from Metrobank. She looks pretty rich, and after buying your drink she says she just made her bonus for the year. So, given the information that the employee made the bonus (and also assuming that her accent is acceptably mid-Atlantic), what is the probability that this employee is from London, and what is the probability that she is from New York? This information is important for your prediction of how events for the evening may unfold…

Fortunately, you can use Bayes' Theorem to work this out. Let *B* be the event that the employee got the bonus; then what you are really looking for is the posterior probabilities $P(A_1 \mid B)$ and $P(A_2 \mid B)$. Looking back to the law of conditional probability, you already know that:

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)}$$

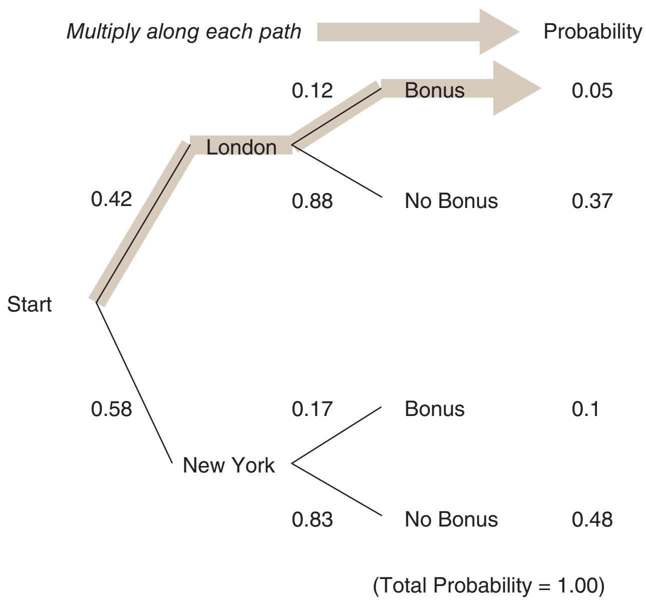

Figure 5.6 Probability Tree for the Human Resource Example

Looking at Figure 5.6, you can also see that:

$$P(A_1 \cap B) = P(A_1)P(B \mid A_1)$$

and to find $P(B)$ you need to consider that it can only happen in two possible ways, either $(A_1 \cap B)$ or $(A_2 \cap B)$, and this means that we can work out that:

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B)$$

Thinking back to the multiplication law, this means that $P(B)$ must also equal:

$$P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)$$

So, if we substitute the multiplication law equations into the conditional probability equation, we get Bayes' Theorem, and we can do the same for $P(A_2 \mid B)$ to get Bayes' Theorem for a two-event case, as shown in Box 5.7.

box 5.7

Bayes' Theorem for Two Events

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)}$$

and:

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)}$$

OK, so let's do it! Using the first equation in Box 5.7, and the probabilities already provided, we get:

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{0.42 \times 0.12}{0.42 \times 0.12 + 0.58 \times 0.17} = \frac{0.0504}{0.0504 + 0.0986}$$

$$= \frac{0.0504}{0.149} = 0.338$$

Then, using the second equation in Box 5.7, we can get:

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{0.58 \times 0.17}{0.42 \times 0.12 + 0.58 \times 0.17} = \frac{0.0986}{0.0516 + 0.0986}$$

$$= \frac{0.0986}{0.1498} = 0.6582$$

Nice work! Now, remember that we started with a probability of 0.58 that the stockbroker was from New York, but given the later information that she got a bonus, the probability can be adjusted to 0.66, with a corresponding 0.34 probability that she is from London. These are our *posterior probabilities*. You can make a joke here about posteriors if you want. I wouldn't of course…

Bayes' Theorem can be used when the events for which we are considering computing posteriors are mutually exclusive (i.e. only one can happen), and the union of the events is all the possible outcomes (i.e. the entire sample space). If this is the case, we are not restricted to two events either, Bayes' Theorem can be extended to any number of posteriors, as shown below in Box 5.8. All you need are the relevant prior probabilities and the conditional probabilities.

box 5.8

Bayes' Theorem

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2) + \ldots + P(A_n)\,P(B \mid A_n)}$$

Now remember that the subscript i refers to any given event (e.g. event 1 or 2 or 3, or whatever), and n refers to the total number of events we are working with (i.e. A_1, A_2, A_3), all the way up to A_n. Go back to Chapter 1 if you are a bit unsure here.

think it over 5.7

Misuse of Conditional Probability

Prosecutor's Fallacy

A murder has been committed. DNA taken from the scene is found to match that of a man's stored sometime earlier in a national database. The man is arrested and charged with the murder. At the trial the prosecution state that the probability is 1 in 1 million that the two samples could have matched by chance. They then go on to claim that therefore the probability that the man is innocent in 1 in 1 million.

This is a false claim.

Let E be the evidence and let I be the event 'the accused is innocent'. Then we are given that $P(E \mid I) = 0.0000001$. But what we want is $P(I \mid E)$.

Defender's Fallacy

In response to the above, the defence might argue that there are 60 million samples in the database, so we would expect on the order of 60 matches. Hence there is a 59 in 60 chance that the accused is innocent. Again though, this is ignoring other evidence, although the exact nature of the fallacy is less clear cut than the prosecutor's example above.

think it over 5.8

False probabilistic reasoning

A judge tells a condemned prisoner that he will be hanged at noon on one weekday in the following week but that the execution will be a surprise to the prisoner. He will not know the day of the hanging until the executioner knocks on his cell door at noon that day. Having reflected on his sentence, the prisoner draws the conclusion that he will escape from the hanging. His reasoning is in several parts. He begins by concluding that the 'surprise hanging' cannot be on a Friday, because if he has not been hanged by Thursday, there is only one day left – and so it will not be a surprise if he is hanged on a Friday. Since the judge's sentence stipulated that the hanging would be a surprise to him, he concludes that it cannot occur on Friday. He then reasons that the surprise hanging cannot be on Thursday either, because Friday has already been eliminated and if he has not been hanged by Wednesday night, the hanging must occur on Thursday, making a Thursday hanging not a surprise either. By similar reasoning he concludes that the hanging can also not occur on Wednesday, Tuesday or Monday. Joyfully he retires to his cell confident that the hanging will not occur at all. The next week, the executioner knocks on the prisoner's door at noon on Wednesday — which, despite all the above, will still be an utter surprise to him. Everything the judge said has come true.

Serious point: You must be very careful when using conditional probability or Bayes' Theorem to make sure you are asking the right question (assigning the probabilities the right way around). For example, you can do some independent research about the case of Sally Clark, who was wrongly convicted on faulty probabilistic reasoning, to see what disastrous effects can occur.

SUMMARY

As I mentioned earlier in this chapter, probability is one of the things I found most difficult to understand when I was first teaching. Some of you I guess will have also found this chapter pretty difficult to get a handle on, but I hope that many of you will have followed along in a step-by-step fashion. If not, perhaps you could go back and try to do this – I promise that doing things step by step is much easier than just rushing through, or, even worse, dipping in and out in a non-sequential fashion.

However, whether or not you followed the techniques and equations, I imagine that many of you will still be wondering 'so what?' Why did I go to all this trouble to introduce probabilities, and where does it fit in with the rest of the material you will learn? Well, the answer to this question is sometimes not too self-evident. After all, I did give you some examples, but is this kind of thing really likely to happen in the real world? Well, actually, yes as you might have seen if you did the independent research suggested earlier about Sally Clark (and she was by no means the only person to be harmed by faulty understanding of probability in the courts). But, in day-to-day life, sometimes it can be hard to see how probability can help you.

So why waste time here? In fact, the concept of probability is the foundation stone of almost all of the later techniques you will learn – whether you will focus on statistical analysis, financial modelling, quality assurance, or virtually any other quantitative analysis technique. Understanding the basic concepts of probability will prove to be essential to understanding later parts of this book. Without it, you will not really make the critical connections about what it is you are actually trying to do with these more advanced techniques.

So it turns out that it is kind of important. Even more so, it offers a very important example to you of things which may not seem that relevant at the time, turning out to be important concepts for later work. If you want to understand how to use quantitative tools correctly, you pretty much have to understand probability. In that respect it is a little like algebra – without it you cannot do many more interesting things, but at the time it seems a little abstract and pointless.

final checklist

You should have covered and understood the following basic content in this chapter:

- ☑ The probability of an event as the likelihood of the occurrence of that event, expressed by a number between 0 (no chance) and 1 (certain).
- ☑ The difference between theoretical and empirical probability.
- ☑ What a probability experiment, a sample space, an outcome and an event are.
- ☑ How a tree diagram can represent multiple step experiments, and how the counting rule can be used to determine the total number of outcomes.
- ☑ How the counting rules regarding permutations and combinations can be used to determine how many permutations and combinations are possible from any given set.
- ☑ How to assign a probability to an outcome using classical, relative frequency and subjective methods.
- ☑ The relationship between the probability of an event and its complement, and the addition law.
- ☑ Conditional probability, joint probability, marginal probability and the multiplication law.
- ☑ Basic Bayesian concepts, and Bayes' Theorem.

EXERCISES

1. You are playing a game which involves tossing a coin. In statistical language the coin is said to be fair (i.e. there is an equal likelihood of a head or a tail). You toss the coin four times and each time it comes up heads.

   a. What is the probability of this occurring?
   b. What is the probability of the next toss coming up tails?

2. You, being an astute business student, can see a way of making money out of this game. You play the game again with a friend but this time after a run of five heads, you ask her to place a bet on whether the next toss would be heads or tails. She bets tails.

   a. Is the following statement true: after a run of heads the chances of a tail must improve?
   b. What is the likelihood of her winning?
   c. What is the probability of the next toss coming up heads?

3. Suppose you are on a game show and you are given the choice of three doors, red, green and blue, from which to choose, one of which has a prize hidden behind it. Suppose you choose the red door. The presenter, who knows where the prize is (and will not choose that door to open), opens the blue door and reveals that there is no prize behind it. He then asks if you wish to change your choice from your initial selection of red.

   a. Will changing your mind at this point improve our chances of winning the prize?
   b. What probability method could you use to determine whether it is worth changing your initial selection?

4. It is lunchtime and you cannot decide what to have. Your friend, who loves stats, buys six cheese sandwiches (which you hate), four ham sandwiches and five egg sandwiches and puts them in a box. You have to put your hand in without being able to see, and pick one. What are the chances of your ending up with:

    a. A cheese sandwich
    b. A ham sandwich
    c. An egg sandwich
    d. Not a cheese sandwich
    e. A cheese or a ham sandwich?

5. Which of the following sets are equal: {1,3,5}, {5,1,3,3}, {1,5,3,1}, {1,3,5,5}?

6. A universal set $U = \{1,2,3,4,5,6,7,8,9\}$ and the sets $A = \{1,2,3,4,5\}$, $B = \{4,5,6,7\}$ and $C = \{5,6,7,8,9\}$. Find:

    a. $A \cup B$
    b. $A \cap B$
    c. $A \cup B$
    d. $A \cap B$

7. Using the same sets as in question 6 find:

    a. $(A \cup B \cap C)$
    b. $C \cup (A \cup B)$
    c. $(C \cup A) \cap (C \cup B)$

8. Use Venn diagrams to illustrate the following:

    a. $A \cup B$
    b. $A \cap \bar{B}$

9. A universal set $U = \{1,2,3,4,5,6,7,8,9\}$ and the sets $A = \{1,2,3,4,5\}$, $B = \{4,5,6,7\}$ and $C = \{5,6,7,8,9\}$. Use Venn diagrams to illustrate the following:

    a. $A \cup B$
    b. $A \cap B$

10. Two events A and B are such that $P(A) = 0.6$ and $P(B) = 0.4$ and $P(A \mid B) = 0.2$.

    a. Using a Venn diagram, determine the probability that neither A nor B occurs. [0.2]
    b. Are the events independent? Give a reason for your answer.
    c. Calculate $P(A \cap B)$

To access additional online resources please visit: **https://study.sagepub.com/leepeters**

The Ballad of Eddie the Easily Distracted: Part 6

Another lecture on probability was leaving Eddie completely cold again. He kind of got the picture of the last lecture, but this one was getting really heavy, with all of the talk about infinity and stuff. But he was glad he had spent a bit of time talking about last week's stuff with Esha in the computer lab – there'd not been much to cover last week, so Eddie took the opportunity to ask a few questions after they'd whizzed through the exercises. Looking at some of the formulas (or 'formulae' as Esha had said, which Eddie thought was a little pretentious to be honest), he was glad he'd paid attention.

Probably the most important part, Eddie thought, was to understand the idea of the continuous distributions, and especially how to use the area under a graph to compute probability. He had a strange tingle of remembrance here, something to do with his calculus classes at school, but he ignored it as he did with most strange tingles.

In fact, after his talk with Esha, Eddie was feeling quite confident about the calculations required. Esha had shown him how, even though they looked intimidating when you looked at the whole thing, they were much easier to do if you broke them down into small parts, based on things you already knew. So the trick seemed to be to learn the earlier steps before trying the later ones.

'It might be boring,' thought Eddie ruefully, 'but the chances of my success approach zero if I don't…'

Esha's Story: Part 6

Esha had a strange feeling about Eddie after the computer lab but couldn't quite locate it somehow. What was it? Oh yes, she was slightly *impressed* that he'd asked her to work through the probability calculations with him after going through the computer lab exercise. In fact, she'd also been somewhat impressed that he'd helped her to do the PC exercise so quickly. Certainly, when he was motivated he could do things pretty quickly! Esha stored that observation away for future reference.

In fact, explaining it had helped Esha too, and she was glad of this because the next part of the course on probability distributions needed some pretty serious thinking. The calculations were reasonably simple, if a little intimidating now and then, but they really were only based on stuff she'd done earlier. The idea of infinity, and continuous distributions where a variable could not take on a single value but only be considered in terms of a range, was hard to get her head around.

She resolved to read the textbook more carefully. On doing so, things became clearer. 'Well,' she thought, 'I suppose infinity is, well, infinite.' Slightly embarrassed by that observation, she continued to think: 'So, every time I think of one number, I can split it into two more, and then two more, and so on, but the key is that I keep doing that *for ever*.' It was beginning to click: 'So then, I could never ever stop splitting, so it is *impossible* to stop at one actual number – there's always a smaller division.' Now she was getting excited, it was starting to make sense, and she knew that once she got this idea under her belt, the rest of it made more sense too. She got up and made a coffee, as she liked to do when she was pleased with her thinking, and then got back to the chapter.