Probability and Random Variables

A Beginner's Guide

David Stirzaker

University of Oxford

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1999

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Times 10/12.5pt, in 3B2 [KT]

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

Library of Congress Cataloguing in Publication data

Stirzaker, David.
Probability and random variables : a beginner's guide / David Stirzaker.
p. cm.
ISBN 0 521 64297 3 (hb)
ISBN 0 521 64445 3 (pb)
1. Probabilities. 2. Random variables. I. Title.
QA273.S75343 1999
519.2–dc21 98-29586 CIP

ISBN 0 521 64297 3 hardback
ISBN 0 521 64445 3 paperback

Contents

1
Introduction

I shot an arrow into the air
It fell to earth, I knew not where.
H.W. Longfellow

O! many a shaft at random sent
Finds mark the archer little meant.
W. Scott

1.1 PREVIEW

This chapter introduces probability as a measure of likelihood, which can be placed on a numerical scale running from 0 to 1. Examples are given to show the range and scope of problems that need probability to describe them. We examine some simple interpretations of probability that are important in its development, and we briefly show how the well-known principles of mathematical modelling enable us to progress. Note that in this chapter exercises and problems are chosen to motivate interest and discussion; they are therefore non-technical, and mathematical answers are not expected.

Prerequisites. This chapter contains next to no mathematics, so there are no prerequisites. Impatient readers keen to get to an equation could proceed directly to chapter 2.

1.2 PROBABILITY

We all know what light is, but it is not easy to tell what it is.
Samuel Johnson

From the moment we first roll a die in a children's board game, or pick a card (*any* card), we start to learn what probability is. But even as adults, it is not easy to *tell* what it is, in the general way.

For mathematicians things are simpler, at least to begin with. We have the following:

Probability is a number between zero and one, inclusive.

This may seem a trifle arbitrary and abrupt, but there are many excellent and plausible reasons for this convention, as we shall show. Consider the following eventualities.

(i) You run a mile in less than 10 seconds.
(ii) You roll two ordinary dice and they show a double six.
(iii) You flip an ordinary coin and it shows heads.
(iv) Your weight is less than 10 tons.

If you think about the relative likelihood (or chance or probability) of these eventualities, you will surely agree that we can compare them as follows.

The chance of running a mile in 10 seconds is *less* than the chance of a double six, which in turn is *less* than the chance of a head, which in turn is *less* than the chance of your weighing under 10 tons. We may write

$$\begin{aligned}\text{chance of 10 second mile} &< \text{chance of a double six}\\ &< \text{chance of a head}\\ &< \text{chance of weighing under 10 tons.}\end{aligned}$$

(Obviously it is assumed that you are reading this on the planet Earth, not on some asteroid, or Jupiter, that you are human, and that the dice are not crooked.)

It is easy to see that we can very often compare probabilities in this way, and so it is natural to represent them on a numerical scale, just as we do with weights, temperatures, earthquakes, and many other natural phenomena. Essentially, this is what numbers are *for*.

Of course, the two extreme eventualities are special cases. It is quite certain that you weigh less than 10 tons; nothing could be more certain. If we represent certainty by unity, then no probabilities exceed this. Likewise it is quite impossible for you to run a mile in 10 seconds or less; nothing could be less likely. If we represent impossibility by zero, then no probability can be less than this. Thus we can, if we wish, present this on a scale, as shown in figure 1.1.

The idea is that any chance eventuality can be represented by a point somewhere on this scale. Everything that is impossible is placed at zero – that the moon is made of

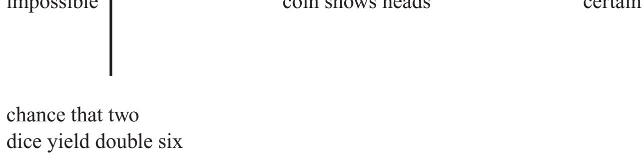

Figure 1.1. A probability scale.

cheese, formation flying by pigs, and so on. Everything that is certain is placed at unity – the moon is not made of cheese, Socrates is mortal, and so forth. Everything else is somewhere in [0, 1], i.e. in the interval between 0 and 1, the more likely things being closer to 1 and the more unlikely things being closer to 0.

Of course, if two things have the same chance of happening, then they are at the same point on the scale. That is what we mean by 'equally likely'. And in everyday discourse everyone, including mathematicians, has used and will use words such as very likely, likely, improbable, and so on. However, any detailed or precise look at probability requires the use of the numerical scale. To see this, you should ponder on just how you would describe a chance that is more than very likely, but less than very very likely.

This still leaves some questions to be answered. For example, the choice of 0 and 1 as the ends of the scale may appear arbitrary, and, in particular, we have not said exactly which numbers represent the chance of a double six, or the chance of a head. We have not even justified the claim that a head is more likely than double six. We discuss all this later in the chapter; it will turn out that if we regard probability as an extension of the idea of proportion, then we can indeed place many probabilities accurately and confidently on this scale.

We conclude with an important point, namely that the chance of a head (or a double six) is just a *chance*. The whole point of probability is to discuss uncertain eventualities *before* they occur. *After* this event, things are completely different. As the simplest illustration of this, note that even though we agree that if we flip a coin and roll two dice then the chance of a head is greater than the chance of a double six, nevertheless it may turn out that the coin shows a tail when the dice show a double six. Likewise, when the weather forecast gives a 90% chance of rain, or even a 99% chance, it may in fact not rain. The chance of a slip on the San Andreas fault this week is very small indeed, nevertheless it may occur today. The antibiotic is overwhelmingly likely to cure your illness, but it may not; and so on.

Exercises for section 1.2

1. Formulate your own definition of probability. Having done so, compare and contrast it with those in appendix I of this chapter.

2. (a) Suppose you flip a coin; there are two possible outcomes, head or tail. Do you agree that the probability of a head is $\frac{1}{2}$? If so, explain why.
(b) Suppose you take a test; there are two possible outcomes, pass or fail. Do you agree that the probability of a pass is $\frac{1}{2}$? If not, explain why not.

3. In the above discussion we claimed that it was intuitively reasonable to say that you are more likely to get a head when flipping a coin than a double six when rolling two dice. Do you agree? If so, explain why.

1.3 THE SCOPE OF PROBABILITY

> . . . nothing between humans is 1 to 3. In fact, I long ago come to the conclusion that all life is 6 to 5 against.
>
> Damon Runyon, *A Nice Price*

Life is a gamble at terrible odds; if it was a bet you wouldn't take it.
Tom Stoppard, *Rosencrantz and Guildenstern are Dead,* Faber and Faber

In the next few sections we are going to spend a lot of time flipping coins, rolling dice, and buying lottery tickets. There are very good reasons for this narrow focus (to begin with), as we shall see, but it is important to stress that probability is of great use and importance in many other circumstances. For example, today seems to be a fairly typical day, and the newspapers contain articles on the following topics (in random order).

1. How are the chances of a child's suffering a genetic disorder affected by a grandparent's having this disorder? And what difference does the sex of child or ancestor make?
2. Does the latest opinion poll reveal the true state of affairs?
3. The lottery result.
4. DNA profiling evidence in a trial.
5. Increased annuity payments possible for heavy smokers.
6. An extremely valuable picture (a Van Gogh) might be a fake.
7. There was a photograph taken using a scanning tunnelling electron microscope.
8. Should risky surgical procedures be permitted?
9. Malaria has a significant chance of causing death; prophylaxis against it carries a risk of dizziness and panic attacks. What do you do?
10. A commodities futures trader lost a huge sum of money.
11. An earthquake occurred, which had not been predicted.
12. Some analysts expected inflation to fall; some expected it to rise.
13. Football pools.
14. Racing results, and tips for the day's races.
15. There is a 10% chance of snow tomorrow.
16. Profits from gambling in the USA are growing faster than any other sector of the economy. (In connection with this item, it should be carefully noted that profits are made by the casino, not the customers.)
17. In the preceding year, British postmen had sustained 5975 dogbites, which was around 16 per day on average, or roughly one every 20 minutes during the time when mail is actually delivered. One postman had sustained 200 bites in 39 years of service.

Now, this list is by no means exhaustive; I could have made it longer. And such a list could be compiled every day (see the exercise at the end of this section). The subjects reported touch on an astonishingly wide range of aspects of life, society, and the natural world. And they all have the common property that chance, uncertainty, likelihood, randomness – call it what you will – is an inescapable component of the story. Conversely, there are few features of life, the universe, or anything, in which chance is not in some way crucial.

Nor is this merely some abstruse academic point; assessing risks and taking chances are inescapable facets of everyday existence. It is a trite maxim to say that life is a lottery; it would be more true to say that life offers a collection of lotteries that we can all, to some extent, choose to enter or avoid. And as the information at our disposal increases, it does not reduce the range of choices but in fact increases them. It is, for example, increasingly difficult successfully to run a business, practise medicine, deal in finance, or engineer things without having a keen appreciation of chance and probability. Of course you can make the attempt, by relying entirely on luck and uninformed guesswork, but in the long run you will probably do worse than someone who plays the odds in an informed way. This is amply confirmed by observation and experience, as well as by mathematics.

Thus, probability is important for all these severely practical reasons. And we have the bonus that it is also entertaining and amusing, as the existence of all those lotteries, casinos, and racecourses more than sufficiently testifies.

Finally, a glance at this and other section headings shows that chance is so powerful and emotive a concept that it is employed by poets, playwrights, and novelists. They clearly expect their readers to grasp jokes, metaphors, and allusions that entail a shared understanding of probability. (This feat has not been accomplished by algebraic structures, or calculus, and is all the more remarkable when one recalls that the *literati* are not otherwise celebrated for their keen numeracy.) Furthermore, such allusions are of very long standing; we may note the comment attributed by Plutarch to Julius Caesar on crossing the Rubicon: ‘Iacta alea est’ (commonly rendered as ‘The die is cast’). And the passage from Ecclesiastes: ‘The race is not always to the swift, or the battle to the strong, but time and chance happen to them all’. The Romans even had deities dedicated to chance, Fors and Fortuna, echoed in Shakespeare’s *Hamlet*: ‘. . . the slings and arrows of outrageous fortune . . .’.

Many other cultures have had such deities, but it is notable that deification has not occurred for any other branch of mathematics. There is no god of algebra.

One recent stanza (by W.H. Henley) is of particular relevance to students of probability, who are often soothed and helped by murmuring it during difficult moments in lectures and textbooks:

> In the fell clutch of circumstance
> I have not winced or cried aloud:
> Under the bludgeonings of chance
> My head is bloody, but unbowed.

Exercise for section 1.3

1. Look at today’s newspapers and mark the articles in which chance is explicitly or implicitly an important feature of the report.

1.4 BASIC IDEAS: THE CLASSICAL CASE

> The perfect die does not lose its usefulness or justification by the fact that real dice fail to live up to it.
>
> W. Feller

Our first task was mentioned above; we need to supply reasons for the use of the standard probability scale, and methods for deciding where various chances should lie on this scale. It is natural that in doing this, and in seeking to understand the concept of probability, we will pay particular attention to the experience and intuition yielded by flipping coins and rolling dice. Of course this is not a very bold or controversial decision;

any theory of probability that failed to describe the behaviour of coins and dice would be widely regarded as useless. And so it would be. For several centuries that we know of, and probably for many centuries before that, flipping a coin (or rolling a die) has been the epitome of probability, the paradigm of randomness. You flip the coin (or roll the die), and nobody can accurately predict how it will fall. Nor can the most powerful computer predict correctly how it will fall, if it is flipped energetically enough.

This is why cards, dice, and other gambling aids crop up so often in literature both directly and as metaphors. No doubt it is also the reason for the (perhaps excessive) popularity of gambling as entertainment. If anyone had any idea what numbers the lottery would show, or where the roulette ball will land, the whole industry would be a dead duck.

At any rate, these long-standing and simple gaming aids do supply intuitively convincing ways of characterizing probability. We discuss several ideas in detail.

I Probability as proportion

Figure 1.2 gives the layout of an American roulette wheel. Suppose such a wheel is spun once; what is the probability that the resulting number has a 7 in it? That is to say, what is the probability that the ball hits 7, 17, or 27? These three numbers comprise a proportion $\frac{3}{38}$ of the available compartments, and so the essential symmetry of the wheel (assuming it is well made) suggests that the required probability ought to be $\frac{3}{38}$. Likewise the

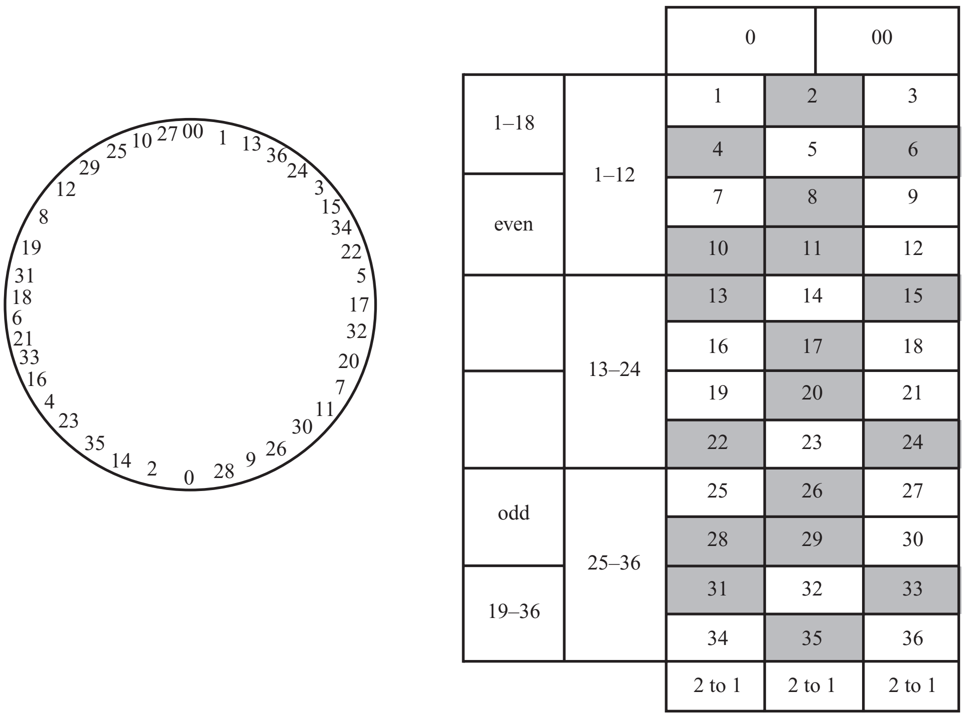

Figure 1.2. American roulette. Shaded numbers are black; the others are red except for the zeros.

probability of an odd compartment is suggested to be $\frac{18}{38} = \frac{9}{19}$, because the proportion of odd numbers on the wheel is $\frac{18}{38}$.

Most people find this proposition intuitively acceptable; it clearly relies on the fundamental symmetry of the wheel, that is, that all numbers are regarded equally by the ball. But this property of symmetry is shared by a great many simple chance activities; it is the same as saying that all possible outcomes of a game or activity are equally likely. For example:

- The ball is equally likely to land in any compartment.
- You are equally likely to select either of two cards.
- The six faces of a die are equally likely to be face up.

With these examples in mind it seems reasonable to adopt the following convention or rule. Suppose some game has n equally likely outcomes, and r of these outcomes correspond to your winning. Then the probability p that you win is r/n. We write

(1) $$p = \frac{r}{n} = \frac{\text{number of ways of winning the game}}{\text{number of possible outcomes of the game}}.$$

This formula looks very simple. Of course, it *is* very simple but it has many useful and important consequences. First note that we always have $0 \leqslant r \leqslant n$, and so it follows that

(2) $$0 \leqslant p \leqslant 1.$$

If $r = 0$, so that it is impossible for you to win, then $p = 0$. Likewise if $r = n$, so that you are certain to win, then $p = 1$. This is all consistent with the probability scale introduced in section 1.2, and supplies some motivation for using it. Furthermore, this interpretation of probability as defined by proportion enables us to place many simple but important chances on the scale.

Example 1.4.1. Flip a coin and choose 'heads'. Then $r = 1$, because you win on the outcome 'heads', and $n = 2$, because the coin shows a head or a tail. Hence the probability that you win, which is also the probability of a head, is $p = \frac{1}{2}$. ○

Example 1.4.2. Roll a die. There are six outcomes, which is to say that $n = 6$. If you win on an even number then $r = 3$, so the probability that an even number is shown is

$$p = \tfrac{3}{6} = \tfrac{1}{2}.$$

Likewise the chance that the die shows a 6 is $\frac{1}{6}$, and so on. ○

Example 1.4.3. Pick a card at random from a pack of 52 cards. What is the probability of an ace? Clearly $n = 52$ and $r = 4$, so that

$$p = \tfrac{4}{52} = \tfrac{1}{13}.$$ ○

Example 1.4.4. A town contains x women and y men; an opinion pollster chooses an adult at random for questioning about toothpaste. What is the chance that the adult is male? Here

$$n = x + y \quad \text{and} \quad r = y.$$

Hence the probability is

$$p = y/(x + y).$$ ○

It may be objected that these results depend on an arbitrary imposition of the ideas of symmetry and proportion, which are clearly not always relevant. Nevertheless, such results and ideas are immensely appealing to our intuition; in fact the first probability calculations in Renaissance Italy take this framework more or less for granted. Thus Cardano (writing around 1520), says of a well-made die: 'One half of the total number of faces always represents equality . . . I can as easily throw 1, 3, or 5 as 2, 4, or 6'.

Here we can clearly see the beginnings of the idea of probability as an expression of proportion, an idea so powerful that it held sway for centuries. However, there is at least one unsatisfactory aspect to this interpretation: it seems that we do not need ever to roll a die to say that the chance of a 6 is $\frac{1}{6}$. Surely actual experiments should have a role in our definitions? This leads to another idea.

II Probability as relative frequency

Figure 1.3 shows the proportion of sixes that appeared in a sequence of rolls of a die. The number of rolls is n, for $n = 0, 1, 2, \ldots$; the number of sixes is $r(n)$, for each n, and the proportion of sixes is

(3) $$p(n) = \frac{r(n)}{n}.$$

What has this to do with the probability that the die shows a six? Our idea of probability as a proportion suggests that the proportion of sixes in n rolls should not be too far from the theoretical chance of a six, and figure 1.3 shows that this seems to be true for large values of n. This is intuitively appealing, and the same effect is observed if you record such proportions in a large number of other repeated chance activities.

We therefore make the following general assertion. Suppose some game is repeated a large number n of times, and in $r(n)$ of these games you win. Then the probability p that

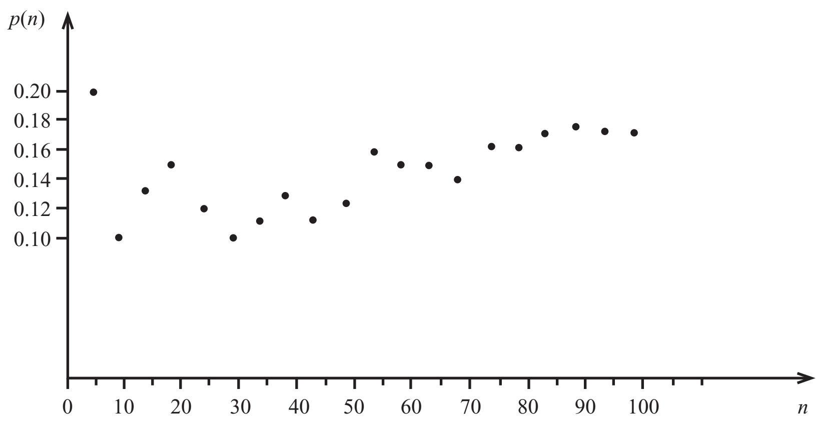

Figure 1.3. The proportion of sixes given in 100 rolls of a die, recorded at intervals of 5 rolls. Figures are from an actual experiment. Of course, $\frac{1}{6} = 0.16\dot{6}$.

you win some future similar repetition of this game is close to $r(n)/n$. We write

$$p \simeq \frac{r(n)}{n} = \frac{\text{number of wins in } n \text{ games}}{\text{number } n \text{ of games}}. \quad (4)$$

The symbol $\simeq$ is read as 'is approximately equal to'. Once again we note that $0 \leqslant r(n) \leqslant n$ and so we may take it that $0 \leqslant p \leqslant 1$.

Furthermore, if a win is impossible then $r(n) = 0$, and $r(n)/n = 0$. Also, if a win is certain then $r(n) = n$, and $r(n)/n = 1$. This is again consistent with the scale introduced in figure 1.1, which is very pleasant. Notice the important point that this interpretation supplies a way of approximately measuring probabilities rather than calculating them merely by an appeal to symmetry.

Since we can now calculate simple probabilities, and measure them approximately, it is tempting to stop there and get straight on with formulating some rules. That would be a mistake, for the idea of proportion gives another useful insight into probability that will turn out to be just as important as the other two, in later work.

III Probability and expected value

Many problems in chance are inextricably linked with numerical outcomes, especially in gambling and finance (where 'numerical outcome' is a euphemism for money). In these cases probability is inextricably linked to 'value', as we now show.

To aid our thinking let us consider an everyday concrete and practical problem. A plutocrat makes the following offer. She will flip a fair coin; if it shows heads she will give you \$1, if it shows tails she will give Jack \$1. What is this offer worth to you? That is to say, for what fair price $\$p$, should you sell it?

Clearly, whatever price $\$p$ this is worth to you, it is worth the same price $\$p$ to Jack, because the coin is fair, i.e. symmetrical (assuming he needs and values money just as much as you do). So, to the pair of you, this offer is altogether worth $\$2p$. But whatever the outcome, the plutocrat has given away \$1. Hence $\$2p = \$1$, so that $p = \frac{1}{2}$ and the offer is worth $\$\frac{1}{2}$ to you.

It seems natural to regard this value $p = \frac{1}{2}$ as a measure of your chance of winning the money. It is thus intuitively reasonable to make the following general rule.

Suppose you receive \$1 with probability p (and otherwise you receive nothing). Then the *value* or fair price of this offer is $\$p$. More generally, if you receive $\$d$ with probability p (and nothing otherwise) then the fair price or expected value of this offer is given by

$$\text{expected value} = pd. \quad (5)$$

This simple idea turns out to be enormously important later on; for the moment we note only that it is certainly consistent with our probability scale introduced in figure 1.1. For example, if the plutocrat definitely gives you \$1 then this is worth exactly \$1 to you, and $p = 1$. Likewise if you are definitely given nothing, then $p = 0$. And it is easy to see that $0 \leqslant p \leqslant 1$, for any such offers.

In particular, for the specific example above we find that the probability of a head when a fair coin is flipped is $\frac{1}{2}$. Likewise a similar argument shows that the probability of a six when a fair die is rolled is $\frac{1}{6}$. (Simply imagine the plutocrat giving \$1 to one of six people selected by the roll of the die.)

The ‘fair price’ of such offers is often called the expected value, or *expectation*, to emphasize its chance nature. We meet this concept again, later on.

We conclude this section with another classical and famous manifestation of probability. It is essentially the same as the first we looked at, but is superficially different.

IV Probability as proportion again

Suppose a small meteorite hits the town football pitch. What is the probability that it lands in the central circle?

Obviously meteorites have no special propensity to hit any particular part of a football pitch; they are equally likely to strike any part. It is therefore intuitively clear that the chance of striking the central circle is given by the proportion of the pitch that it occupies. In general, if $|A|$ is the area of the pitch in which the meteorite lands, and $|C|$ is the area of some part of the pitch, then the probability p that C is struck is given by $p = |C|/|A|$.

Once again we formulate a general version of this as follows. Suppose a region A of the plane has area $|A|$, and C is some part of A with area $|C|$. If a point is picked at random in A, then the probability p that it lies in C is given by

$$p = \frac{|C|}{|A|}. \tag{6}$$

As before we can easily see that $0 \leqslant p \leqslant 1$, where $p = 0$ if C is empty and $p = 1$ if $C = A$.

Example 1.4.5. An archery target is a circle of radius 2. The bullseye is a circle of radius 1. A naive archer is equally likely to hit any part of the target (if she hits it at all) and so the probability of a bullseye for an arrow that hits the target is

$$p = \frac{\text{area of bullseye}}{\text{area of target}} = \frac{\pi \times 1^2}{\pi \times 2^2} = \frac{1}{4}. \qquad \bigcirc$$

Exercises for section 1.4

1. Suppose you read in a newspaper that the proportion of \$20 bills that are forgeries is 5%. If you possess what appears to be a \$20 bill, what is its expected value? Could it be more than \$19? Or could it be less? Explain! (Does it make any difference how you acquired the bill?)
2. A point P is picked at random in the square $ABCD$, with sides of length 1. What is the probability that the distance from P to the diagonal AC is less than $\frac{1}{6}$?

1.5 BASIC IDEAS; THE GENERAL CASE

> We must believe in chance, for how else can we account for the successes of those we detest?
>
> Anon.

We noted that a theory of probability would be hailed as useless if it failed to describe the behaviour of coins and dice. But of course it would be equally useless if it failed to describe anything else, and moreover many real dice and coins (especially dice) have been known to be biased and asymmetrical. We therefore turn to the question of assigning probabilities in activities that do not necessarily have equally likely outcomes.

It is interesting to note that the desirability of doing this was implicitly recognized by Cardano (mentioned in the previous section) around 1520. In his *Book on Games of Chance*, which deals with supposedly fair dice, he notes that

> 'Every die, even if it is acceptable, has its favoured side.'

However, the ideas necessary to describe the behaviour of such biased dice had to wait for Pascal in 1654, and later workers. We examine the basic notions in turn; as in the previous section, these notions rely on our concept of probability as an extension of proportion.

I Probability as relative frequency

Once again we choose a simple example to illustrate the ideas, and a popular choice is the pin, or tack. Figure 1.4 shows a pin, called a Bernoulli pin. If such a pin is dropped onto a table the result is a success, S, if the point is not upwards; otherwise it is a failure, F.

What is the probability p of success? Obviously symmetry can play no part in fixing p, and Figure 1.5, which shows more Bernoulli pins, indicates that mechanical arguments will not provide the answer.

The only course of action is to drop many similar pins (or the same pin many times), and record the proportion that are successes (point down). Then if n are dropped, and $r(n)$ are successes, we anticipate that the long-run proportion of successes is near to p, that is:

(1) $$p \simeq \frac{r(n)}{n}, \quad \text{for large } n.$$

Figure 1.4. A Bernoulli pin.

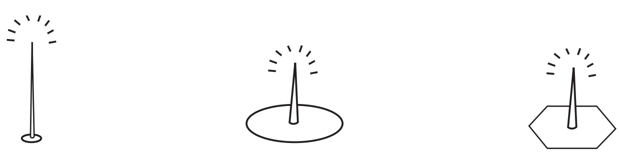

Figure 1.5. More Bernoulli pins.

If you actually obtain a pin and perform this experiment, you will get a graph like that of figure 1.6. It does seem from the figure that $r(n)/n$ is settling down around some number p, which we naturally interpret as the probability of success. It may be objected that the ratio changes every time we drop another pin, and so we will never obtain an exact value for p. But this gap between the real world and our descriptions of it is observed in all subjects at all levels. For example, geometry tells us that the diagonal of a unit square has length $\sqrt{2}$. But, as A. A. Markov has observed,

> If we wished to verify this fact by measurements, we should find that the ratio of diagonal to side is different for different squares, and is never $\sqrt{2}$.

It may be regretted that we have only this somewhat hit-or-miss method of measuring probability, but we do not really have any choice in the matter. Can you think of any other way of estimating the chance that the pin will fall point down? And even if you did think of such a method of estimation, how would you decide whether it gave the right answer, except by flipping the pin often enough to see? We can illustrate this point by considering a basic and famous example.

Example 1.5.1: sex ratio. What is the probability that the next infant to be born in your local hospital will be male? Throughout most of the history of the human race it was taken for granted that essentially equal numbers of boys and girls are born (with some fluctuations, naturally). This question would therefore have drawn the answer $\frac{1}{2}$, until recently.

However, in the middle of the 16th century, English parish churches began to keep fairly detailed records of births, marriages, and deaths. Then, in the middle of the 17th century, one John Graunt (a draper) took the trouble to read, collate, and tabulate the numbers in various categories. In particular he tabulated the number of boys and girls whose births were recorded in London in each of 30 separate years.

To his, and everyone else's, surprise, he found that in every single year more boys were born than girls. And, even more remarkably, the ratio of boys to girls varied very little between these years. In every year the ratio of boys to girls was close to 14:13. The meaning and significance of this unarguable truth inspired a heated debate at the time. For us, it shows that the probability that the next infant born will be male, is approximately $\frac{14}{27}$. A few moments thought will show that there is no other way of answering the general question, other than by finding this relative frequency.

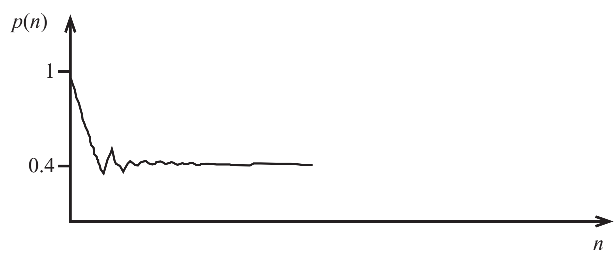

Figure 1.6. Sketch of the proportion $p(n)$ of successes when a Bernoulli pin is dropped n times. For this particular pin, p seems to be settling down at approximately 0.4.

It is important to note that the empirical frequency differs from place to place and from time to time. Graunt also looked at the births in Romsey over 90 years and found the empirical frequency to be 16:15. It is currently just under 0.513 in the USA, slightly less than $\frac{14}{27}$ ($\simeq 0.519$) and $\frac{16}{31}$ ($\simeq 0.516$).

Clearly the idea of probability as a relative frequency is very attractive and useful. Indeed it is generally the only interpretation offered in textbooks. Nevertheless, it is not always enough, as we now discuss.

II Probability as expected value

The problem is that to interpret probability as a relative frequency requires that we can repeat some game or activity as many times as we wish. Often this is clearly not the case. For example, suppose you have a Russian Imperial Bond, or a share in a company that is bankrupt and is being liquidated, or an option on the future of the price of gold. What is the probability that the bond will be redeemed, the share will be repaid, or the option will yield a profit? In these cases the idea of expected value supplies the answer. (For simplicity, we assume constant money values and no interest.)

The ideas and argument are essentially the same as those that we used in considering the benevolent plutocrat in section 1.4, leading to equation (5) in that section. For variety, we rephrase those notions in terms of simple markets. However, a word of warning is appropriate at this point. Real markets are much more complicated than this, and what we call the fair price or expected value will not usually be the actual or agreed market price in any case, or even very close to it. This is especially marked in the case of deals which run into the future, such as call options, put options, and other complicated financial derivatives. If you were to offer prices based on fairness or expected value as discussed here and above, you would be courting total disaster, or worse. See the discussion of bookmakers' odds in section 2.12 for further illustration and words of caution.

Suppose you have a bond with face value \$1, and the probability of its being redeemed at par (that is, for \$1) is p. Then, by the argument we used in section 1.4, the expected value μ, or fair price, of this bond is given by $\mu = p$. More generally, if the bond has face value \$$d$ then the fair price is \$$dp$.

Now, as it happens, there are markets in all these things: you can buy Imperial Chinese bonds, South American Railway shares, pork belly futures, and so on. It follows that if the market gives a price μ for a bond with face value d, then it gives the probability of redemption as roughly

$$p = \frac{\mu}{d}. \tag{2}$$

Example 1.5.2. If a bond for a million roubles is offered to you for one rouble, and the sellers are assumed to be rational, then they clearly think the chance of the bond's being bought back at par is less than one in a million. If you buy it, then presumably you believe the chances are more than one in a million. If you thought the chances were less, you would reduce your offer. If you both agree that one rouble is a fair price for the bond, then you have assigned the value $p = 10^{-6}$ for the probability of its redemption. Of course this may vary according to various rumours and signals from the relevant banks and government (and note that the more ornate and attractive bonds now have some intrinsic value, independent of their chance of redemption). ○

This example leads naturally to our final candidate for an interpretation of probability.

III Probability as an opinion or judgement

In the previous example we were able to assign a probability because the bond had an agreed fair price, *even though* this price was essentially a matter of opinion. What happens if we are dealing with probabilities that are purely personal opinions? For example, what is the probability that a given political party will win the next election? What is the probability that small green aliens regularly visit this planet? What is the probability that some accused person is guilty? What is the probability that a given, opaque, small, brick building contains a pig?

In each of these cases we could perhaps obtain an estimate of the probability by persuading a bookmaker to compile a number of wagers and so determine a fair price. But we would be at a loss if nobody were prepared to enter this game. And it would seem to be at best a very artificial procedure, and at worst extremely inappropriate, or even illegal. Furthermore, the last resort, betting with yourself, seems strangely unattractive.

Despite these problems, this idea of probability as a matter of opinion is often useful, though we shall not use it in this text.

Exercises for section 1.5

1. A picture would be worth \$1000 000 if genuine, but nothing if a fake. Half the experts say it's a fake, half say it's genuine. What is it worth? Does it make any difference if one of the experts is a millionaire?

2. A machine accepts dollar bills and sells a drink for \$1. The price is raised to 120c. Converting the machine to accept coins or give change is expensive, so it is suggested that a simple randomizer is added, so that each customer who inserts \$1 gets nothing with probability $1/6$, or the can with probability $5/6$, and that this would be fair because the expected value of the output is $120 \times 5/6 = 100\text{c} = \$1$, which is exactly what the customer paid. Is it indeed fair?

In the light of this, discuss how far our idea of a fair price depends on a surreptitious use of the concept of repeated experiments.

Would you buy a drink from the modified machine?

1.6 MODELLING

> If I wish to know the chances of getting a complete hand of 13 spades, I do not set about dealing hands. It would take the population of the world billions of years to obtain even a bad estimate of this.
>
> John Venn

The point of the above quote is that we need a theory of probability to answer even the simplest of practical questions. Such theories are called *models*.

Example 1.6.1: cards. For the question above, the usual model is as follows. We assume that all possible hands of cards are equally likely, so that if the number of all possible hands is n, then the required probability is n^{-1}. ○

Experience seems to suggest that for a well-made, well-shuffled pack of cards, this answer is indeed a good guide to your chances of getting a hand of spades. (Though we must remember that such complete hands occur more often than this predicts, because humorists stack the pack, as a 'joke'.) Even this very simple example illustrates the following important points very clearly.

First, the model deals with abstract things. We cannot *really* have a perfectly shuffled pack of perfect cards; this 'collection of equally likely hands' is actually a fiction. We create the idea, and then use the rules of arithmetic to calculate the required chances. This is characteristic of all mathematics, which concerns itself only with rules defining the behaviour of entities which are themselves undefined (such as 'numbers' or 'points').

Second, the use of the model is determined by our interpretation of the rules and results. We do not need an interpretation of what chance is to calculate probabilities, but without such an interpretation it is rather pointless to do it.

Similarly, you do not need to have an interpretation of what lines and points are to do geometry and trigonometry, but it would all be rather pointless if you did not have one. Likewise chess is just a set of rules, but if checkmate were not interpreted as victory, not many people would play.

Use of the term 'model' makes it easier to keep in mind this distinction between theory and reality. By its very nature a model cannot include all the details of the reality it seeks to represent, for then it would be just as hard to comprehend and describe as the reality we want to model. At best, our model should give a reasonable picture of some small part of reality. It has to be a simple (even crude) description; and we must always be ready to scrap or improve a model if it fails in this task of accurate depiction. That having been said, old models are often still useful. The theory of relativity supersedes the Newtonian model, but all engineers use Newtonian mechanics when building bridges or motor cars, or probing the solar system.

This process of observation, model building, analysis, evaluation, and modification is called *modelling*, and it can be conveniently represented by a diagram; see figure 1.7. (This diagram is therefore in itself a model; it is a model for the modelling process.)

In figure 1.7, the top two boxes are embedded in the real world and the bottom two boxes are in the world of models. Box A represents our observations and experience of some phenomenon, together with relevant knowledge of related events and perhaps past experience of modelling. Using this we construct the rules of a model, represented by box B. We then use the techniques of logical reasoning, or mathematics, to deduce the way in which the model will behave. These properties of the model can be called theorems; this stage is represented by box C. Next, these characteristics of the model are interpreted in terms of predictions of the way the corresponding real system should work, denoted by box D. Finally, we perform appropriate experiments to discover whether these predictions agree with observation. If they do not, we change or scrap the model and go round the loop again. If they do, we hail the model as an engine of discovery, and keep using it to make predictions – until it wears out or breaks down. This last step is called using or checking the model or, more grandly, validation.

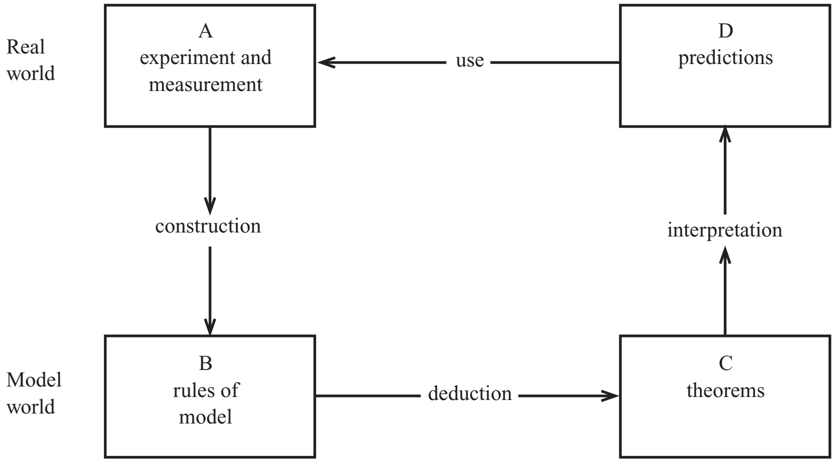

Figure 1.7. A model for modelling.

This procedure is so commonplace that we rather take it for granted. For example, it has been used every time you see a weather forecast. Meteorologists have observed the climate for many years. They have deduced certain simple rules for the behaviour of jet streams, anticyclones, occluded fronts, and so on. These rules form the model. Given any configuration of airflows, temperatures, and pressures, the rules are used to make a prediction; this is the weather forecast. Every forecast is checked against the actual outcome, and this experience is used to improve the model.

Models form extraordinarily powerful and economical ways of thinking about the world. In fact they are often so good that the model is confused with reality. If you ever think about atoms, you probably imagine little billiard balls; more sophisticated readers may imagine little orbital systems of elementary particles. Of course atoms are not 'really' like that; these visions are just convenient old models.

We illustrate the techniques of modelling with two simple examples from probability.

Example 1.6.2: setting up a lottery. If you are organizing a lottery you have to decide how to allocate the prize money to the holders of winning tickets. It would help you to know the chances of any number winning and the likely number of winners. Is this possible? Let us consider a specific example.

Several national lotteries allow any entrant to select six numbers in advance from the integers 1 to 49 inclusive. A machine then selects six balls at random (without replacement) from an urn containing 49 balls bearing these numbers. The first prize is divided among entrants selecting these numbers.

Because of the nature of the apparatus, it seems natural to assume that any selection of six numbers is equally likely to be drawn. Of course this assumption is a mathematical model, not a physical law established by experiment. Since there are approximately 14 million different possible selections (we show this in chapter 3), the model predicts that your chance, with one entry, of sharing the first prize is about one in 14 million. Figure 1.8 shows the relative frequency of the numbers drawn in the first 1200 draws. It does not seem to discredit or invalidate the model so far as one can tell.

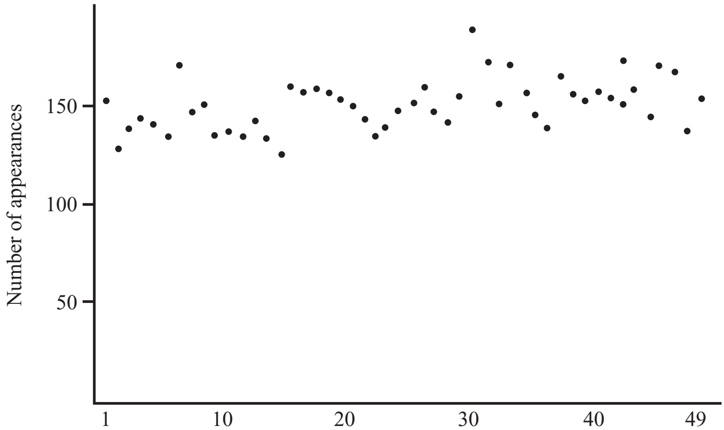

Figure 1.8. Frequency plot of an actual 6–49 lottery after 1200 drawings. The numbers do seem equally likely to be drawn.

The next question you need to answer is, how many of the entrants are likely to share the first prize? As we shall see, we need in turn to ask, how do lottery entrants choose their numbers?

This is clearly a rather different problem; unlike the apparatus for choosing numbers, gamblers choose numbers for various reasons. Very few choose at random; they use birthdays, ages, patterns, and so on. However, you might suppose that for any gambler chosen at random, that choice of numbers would be evenly distributed over the possibilities.

In fact this model would be wrong; when the actual choices of lottery numbers are examined, it is found that in the long run the chances that the various numbers will occur are very far from equal; see figure 1.9. This clustering of preferences arises because people choose numbers in lines and patterns which favour central squares, and they also favour the top of the card. Data like this would provide a model for the distribution of likely payouts to winners. ○

It is important to note that these remarks do not apply only to lotteries, cards, and dice. Venn's observation about card hands applies equally well to almost every other aspect of life. If you wished to design a telephone exchange (for example), you would first of all construct some mathematical models that could be tested (you would do this by making assumptions about how calls would arrive, and how they would be dealt with). You can construct and improve any number of mathematical models of an exchange very cheaply. Building a faulty real exchange is an extremely costly error.

Likewise, if you wished to test an aeroplane to the limits of its performance, you would be well advised to test mathematical models first. Testing a real aeroplane to destruction is somewhat risky.

So we see that, in particular, models and theories can save lives and money. Here is another practical example.

1	2	3	4	**5**
6	**7**	**8**	**9**	10
11	**12**	**13**	**14**	15
16	**17**	**18**	**19**	20
21	**22**	**23**	**24**	25
26	**27**	**28**	**29**	30
31	32	**33**	34	*35*
36	37	38	39	*40*
41	42	43	44	*45*
46	*47*	*48*	*49*	

Figure 1.9. Popular and unpopular lottery numbers: **bold**, most popular; roman, intermediate popularity; *italic*, least popular.

Example 1.6.3: first significant digit. Suppose someone offered the following wager:

(i) select any large book of numerical tables (such as a census, some company accounts, or an almanac);
(ii) pick a number from this book at random (by any means);
(iii) if the first significant digit of this number is one of $\{5, 6, 7, 8, 9\}$, then you win \$1; if it is one of $\{1, 2, 3, 4\}$, you lose \$1.

Would you accept this bet? You might be tempted to argue as follows: a reasonable intuitive model for the relative chances of each digit is that they are equally likely. On this model the probability p of winning is $\frac{5}{9}$, which is greater than $\frac{1}{2}$ (the odds on winning would be $5:4$), so it seems like a good bet. However, if you do some research and actually pick a large number of such numbers at random, you will find that the relative frequencies of each of the nine possible first significant digits are given approximately by

$$\begin{aligned} f_1 &= 0.301, & f_2 &= 0.176, & f_3 &= 0.125, \\ f_4 &= 0.097, & f_5 &= 0.079, & f_6 &= 0.067, \\ f_7 &= 0.058, & f_8 &= 0.051, & f_9 &= 0.046. \end{aligned}$$

Thus empirically the chance of your winning is

$$f_5 + f_6 + f_7 + f_8 + f_9 = 0.3$$

The wager offered is not so good for you! (Of course it would be quite improper for a mathematician to win money from the ignorant by this means.) This empirical distribution is known as Benford's law, though we should note that it was first recorded by S. Newcomb (a good example of Stigler's law of eponymy). ○

We see that intuition is necessary and helpful in constructing models, but not sufficient; you also need experience and observations. A famous example of this arose in particle physics. At first it was assumed that photons and protons would satisfy the same statistical rules, and models were constructed accordingly. Experience and observations showed that in fact they behave differently, and the models were revised.

The theory of probability exhibits a very similar history and development, and we approach it in similar ways. That is to say, we shall construct a model that reflects our experience of, and intuitive feelings about, probability. We shall then deduce results and make predictions about things that have either not been explained or not been observed, or both. These are often surprising and even counter intuitive. However, when the predictions are tested against experiment they are almost always found to be good. Where they are not, new theories must be constructed.

It may perhaps seem paradoxical that we can explore reality most effectively by playing with models, but this fact is perfectly well known to all children.

Exercise for section 1.6

1. Discuss how the development of the high-speed computer is changing the force of Venn's observation, which introduced this section.

1.7 MATHEMATICAL MODELLING

> There are very few things which we know, which are not capable of being reduced to a mathematical reasoning; and when they cannot, it is a sign our knowledge of them is very small and confused; and where a mathematical reasoning can be had, it is as great a folly to make use of any other, as to grope for a thing in the dark, when you have a candle standing by you.
>
> John Arbuthnot, *Of the Laws of Chance*

The quotation above is from the preface to the first textbook on probability to appear in English. (It is in a large part a translation of a book by Huygens, which had previously appeared in Latin and Dutch.) Three centuries later, we find that mathematical reasoning is indeed widely used in all walks of life, but still perhaps not as much as it should be. A small survey of the reasons for using mathematical methods would not be out of place. The first question is, why be abstract at all? The blunt answer is that we have no choice, for many reasons.

In the first place, as several examples have made clear, practical probability is inescapably numerical. Betting odds can only be numerical, monetary payoffs are numerical, stock exchanges and insurance companies float on a sea of numbers. And even the simplest and most elementary problems in bridge and poker, or in lotteries, involve counting things. And this counting is often not a trivial task.

Second, the range of applications demands abstraction. For example, consider the following list of real activities:

- customers in line at a post office counter
- cars at a toll booth
- data in an active computer memory

- a pile of cans in a supermarket
- telephone calls arriving at an exchange
- patients arriving at a trauma clinic
- letters in a mail box

All these entail 'things' or 'entities' in one or another kind of 'waiting' state, before some 'action' is taken. Obviously this list could be extended indefinitely. It is desirable to abstract the essential structure of all these problems, so that the results can be interpreted in the context of whatever application happens to be of interest. For the examples above, this leads to a model called the theory of queues.

Third, we may wish to discuss the behaviour of the system without assigning specific numerical values to the rate of arrival of the objects (or customers), or to the rate at which they are processed (or serviced). We may not even know these values. We may wish to examine the way in which congestion depends generally on these rates. For all these reasons we are naturally forced to use all the mathematical apparatus of symbolism, logic, algebra, and functions. This is in fact very good news, and these methods have the simple practical and mechanical advantage of making our work very compact. This alone would be sufficient! We conclude this section with two quotations chosen to motivate the reader even more enthusiastically to the advantages of mathematical modelling. They illustrate the fact that there is also a considerable gain in understanding of complicated ideas if they are simply expressed in concise notation. Here is a definition of commerce.

> Commerce: a kind of transaction, in which A plunders from B the goods of C, and for compensation B picks the pocket of D of money belonging to E.
>
> Ambrose Bierce, *The Devil's Dictionary*

The whole pith and point of the joke evaporates completely if you expand this from its symbolic form. And think of the expansion of effort required to write it. Using algebra is the reason – or at least one of the reasons – why mathematicians so rarely get writer's cramp or repetitive strain injury.

We leave the final words on this matter to Abraham de Moivre, who wrote the second textbook on probability to appear in English. It first appeared in 1717. (The second edition was published in 1738 and the third edition in 1756, posthumously, de Moivre having died on 27 November, 1754 at the age of 87.) He says in the preface:

> Another use to be made of this Doctrine of Chances is, that it may serve in conjunction with the other parts of mathematics as a fit introduction to the art of reasoning; it being known by experience that nothing can contribute more to the attaining of that art, than the consideration of a long train of consequences, rightly deduced from undoubted principles, of which this book affords many examples. To this may be added, that some of the problems about chance having a great appearance of simplicity, the mind is easily drawn into a belief, that their solution may be attained by the mere strength of natural good sense; which generally proving otherwise, and the mistakes occasioned thereby being not infrequent, it is presumed that a book of this kind, which teaches to distinguish truth from what seems so nearly to resemble it, will be looked on as a help to good reasoning.

These remarks remain as true today as when de Moivre wrote them around 1717.