FROM COSMOS TO CHAOS

The Science of Unpredictability

Peter Coles

OXFORD
UNIVERSITY PRESS

Contents

2

The Logic of Uncertainty

The theory of probabilities . . . is only common sense reduced to calculus.

Pierre Simon, Marquis de Laplace, A Philosophical Essay on Probabilities

First Principles

Since the subject of this book is probability, its meaning and its relevance for science and society, I am going to start in this chapter with a short explanation of how to go about the business of calculating probabilities for some simple examples. I realize that this is not going to be easy. I have from time to time been involved in teaching the laws of probability to high school and university students, and even the most mathematically competent often find it very difficult to get the hang of it. The difficulty stems not from there being lots of complicated rules to learn, but from the fact that there are so few. In the field of probability it is not possible to proceed by memorizing worked solutions to well known (if sometimes complex) problems, which is how many students approach mathematics. The only way forward is to *think*. That is why it is difficult, and also why it is fun.

I will start by dodging the issue of what probability actually *means* and concentrate on how to use it. The controversy surrounding the interpretation of such a common word is the principal subject of Chapter 4, and crops up throughout the later chapters too. What we can say for sure is that a probability is a number that lies between 0 and 1. The two limits are intuitively obvious. An event with zero probability is something that just cannot happen. It must be logically or physically impossible. An event with unit probability is certain. It must happen, and the converse is logically or physically impossible. In between 0 and 1 lies the crux. You have some idea of what it means to say, for example, that the probability of a fair coin landing heads-up is one-half, or that the probability of a fair dice showing a 6 when you roll it is 1/6. Your understanding of these statements (and others like them) is likely to fall in one or other of the following two basic categories. Either the probability represents what will happen if you toss the coin a large number of times, so that it represents some kind of frequency in a long run of repeated trials, or it is some measure of your assessment of the symmetry (or lack of it) in the situation and your subsequent inability to distinguish possible outcomes. A fair dice has six faces; they all look the same, so there is no reason why any one face should have a higher probability of coming up than any other. The probability of a 6 should therefore be the same as any other face. There are six faces, so the required answer must be 1/6. Whichever way you like to think of probability does not really matter for the purposes of this elementary introduction, so just use whichever you feel comfortable with, at least for the time being. The hard sell comes later.

To keep things as simple as possible, I am going to use examples from familiar games of chance. The simplest involving coin-tossing, rolls of a dice, drawing balls from an urn, and standard packs of playing cards. These are the situations for which the mathematical theory of probability was originally developed, so I am really just following history in doing this.

Let us start by defining an *event* to be some outcome of a 'random' experiment. In this context, 'random' means that we do not know how to predict the outcome with certainty. The toss of a coin is governed by Newtonian mechanics, so in principle, we should be able to predict it. However, the coin is usually spun quickly, with no attention given to its initial direction, so that we just accept the outcome will be randomly either head or tails. I have never managed to get a coin to land on its edge, so we will ignore that possibility. In the toss of a coin, there are two possible outcomes of the experiment, so our event may be either of these. Event A might be that 'the coin shows heads'. Event B might be that 'the coin shows tails'. These are the only two possibilities and they are *mutually exclusive* (they cannot happen at the same time). These two events are also *exhaustive*, in that they represent the entire range of possible outcomes of the experiment. We might as well say, therefore, that the event B is the same as 'not A', which we can denote A^*. Our first basic rule of probability is that

$$P(A) + P(A^*) = 1,$$

which basically means that we can be certain that either something (A) happens or it does not (A^*). We can generalize this to the case where we have several mutually exclusive and exhaustive events: A, B, C, and so on. In this case the sum of all probabilities must be 1: however many outcomes are possible, one and only one of them has to happen.

$$P(A) + P(B) + P(C) + \cdots = 1,$$

This is taking us towards the rule for combining probabilities using the operation 'OR'. If two events A and B are mutually exclusive then the probability of *either A or B* is usually written $P(A \cup B)$. This can be obtained by adding the probabilities of the respective events, that is,

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

However, this is not the whole story because not all events are mutually exclusive. The general rule for combining probabilities like this will have to wait a little.

In the coin-tossing example, the event we are interested in is simply one of the outcomes of the experiment ('heads' or 'tails'). In a throw of a dice, a similar type of event A might be that the score is a 6. However, we might instead ask for the probability that the roll of a dice produces an even number. How do we assign a probability for this? The answer is to reduce everything to the elementary outcomes of the experiment which, by reasons of symmetry or ignorance (or both), we can assume to have equal probability. In the roll of a dice, the six individual faces are taken to be equally probable. Each of these must be assigned a probability of 1/6, so the probability of getting a six must also be 1/6. The probability of getting any even number is found by calculating which of the elementary outcomes lead to this composite event and then adding them together. The possible scores are 1, 2, 3, 4, 5, or 6. Of these 2, 4, and 6 are even. The probability of an even number is therefore given by $P(\text{even}) = P(2) + P(4) + P(6) = 1/2$. There is, of course a quicker way to get this answer. Half the possible throws are even, so the probability must be 1/2. You could imagine the faces of the dice were coloured red if odd and black if even. The probability of a black face coming up would be 1/2. There are various tricks like this that can be deployed to calculate complicated probabilities.

In the language of gambling, probabilities are often expressed in terms of odds. If an event has probability p then the odds on it happening are expressed as the ratio $p : (1 - p)$, after some appropriate cancellation. If $p = 0.5$ then the odds are 1 : 1 and we have an even money bet. If the probability is 1 : 3 then the odds are 1/3 : 2/3, or after cancelling the threes, 2 : 1 against. The process of enumerating all the possible elementary outcomes of an experiment can be quite laborious, but it is by far the safest way to calculate odds.

Now let us complicate things a little further with some examples using playing cards. For those of you who did not misspend your youth playing with cards like I did, I should remind you that a standard pack of playing cards has 52 cards. There are 4 suits: clubs (♣), diamonds (♦), hearts (♥) and spades (♠). Clubs and spades are coloured black, while diamonds and hearts are red. Each suit contains thirteen cards, including an Ace (A), the plain numbered cards (2, 3, 4, 5, 6, 7, 8, 9, and 10), and the face cards: Jack (J), Queen (Q), and King (K). In most games the most valuable is the Ace, following by King, Queen, and Jack and then from 10 down to 2.

Suppose we shuffle the cards and deal one. Shuffling is taken to mean that we have lost track of where all the cards are in the pack, and consequently each one is equally likely to be dealt. Clearly the elementary outcomes number 52 in total, each one being a particular card. Each of these has probability 1/52. Let us try some simple examples of calculating combined probabilities.

What is the probability of a red card being dealt? There are a number of ways of doing this, but I will use the brute-force way first. There are 52 cards. The red ones are diamonds or heart suits, each of which has 13 cards. There are therefore 26 red cards, so the probability is 26 lots of 1/52, or one-half. The simplest alternative method is to say there are only two possible colours and each colour applies to the same number of cards. The probability therefore must be 1/2.

What is the probability of dealing a king? There are 4 kings in the pack and 52 cards in total. The probability must be $4/52 = 1/13$. Alternatively there are four suits with the same type of cards. Since we do not care about the suit, the probability of getting a king is the same as if there were just one suit of 13 cards, one of which is a king. This again gives 1/13 for the answer.

What is the probability that the card is a red jack or a black queen? How many red jacks are there? Only two: J♦ and J♥. How many black queens are there? Two: Q♣ and Q♠. The required probability is therefore 4/52, or 1/13 again.

What is the probability that the card we pull out is either a red card or a seven? This is more difficult than the previous examples, because it requires us to build a more complicated combination of outcomes. How many sevens are there? There are four, one of each suit. How many red cards are there? Well, half the cards are red so the answer to that question is 26. But two of the sevens are themselves red so these two events are not mutually exclusive. What do we do?

This brings us to the general rules for combining probabilities whether or not we have exclusivity. The general rule for combining with 'or' is

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

The extra bit that has appeared compared to the previous version, $P(A \cap B)$, is the probability of A *and* B both being the case. This formula is illustrated in the figure using a Venn diagram. If you just add the probabilities of events A and B then the intersection (if it exists) is counted twice. It must be subtracted off to get the right answer, hence the result I quoted above.

To see how this formula works in practice, let us calculate the separate components separately in the example I just discussed. First we can directly work out the left-hand side by enumerating the required probabilities. Each card is mutually exclusive of any other, so we can do this straightforwardly. Which cards satisfy the requirement of redness or seven-ness? Well, there are four sevens for a start. There are then two entire suits of red cards, numbering 26 altogether. But two of these 26 are red sevens (7♦ and 7♥) and I have already counted those. Writing all the possible cards down and crossing out the two duplicates leaves 28: two red suits plus two black sevens. The answer for the probability is therefore 28/52 which is 7/13.

Now let us look at the right-hand side. Let A be the event that the card is a seven and B be the event that it is a red card. There are four sevens, so $P(A) = 4/52 = 1/13$. There are 26 red cards, so $P(B) = 26/52 = 1/2$. What we need to know is $P(A \cap B)$, in other words how many of the 52 cards are both red and sevens? The answer is 2, the 7♦ and 7♠, so this probability is $2/52 = 1/26$. The right-hand side therefore becomes $1/13 + 1/2 - 1/26$, which is the same answer as before.

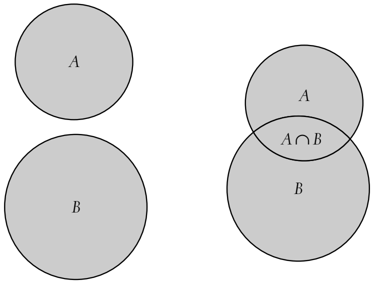

Figure 1 Venn diagrams and probabilities. On the left the two sets A and B are disjoint, so the probability of their intersection is zero. The probability of A or B, P(A∪B) is then just P(A)+P(B). On the right the two sets do intersect so P(A∪B) is given by P(A)+P(B)-P(A∩B).

There is a general formula for the construction of the 'and' probability $P(A \cap B)$, which together with the 'or' formula, is basically all there is to probability theory. The required form is

$$P(A \cap B) = P(A)P(B|A).$$

This tells us the joint probability of two events A and B in terms of the probability of one of them $P(A)$ multiplied by the conditional probability of the second event given the first, $P(B \mid A)$. Conditioning probabilities are probably the most difficult bit of this whole story, and in my experience they are where most people go wrong when trying to do calculations. Forgive me if I labour this point in the following.

The first thing to say about the conditional probability $P(B \mid A)$ is that it need not be the same as $P(B)$. Think of the entire set of possible outcomes of an experiment. In general, only some of these outcomes may be consistent with the event A. If you condition on the event A having taken place then the space of possible outcomes consequently shrinks, and the probability of B in this reduced space may not be the same as it was before the event A was imposed. To see this, let us go back to our example of the red cards and the sevens. Assume that we have picked a red card. The space of possibilities has now shrunk to 26 of the original 52 outcomes. The probability that we have a seven is now just 2 out of 26, or 1/13. In this case $P(A) = 1/2$ for getting a red card, times 1/13 for the conditional probability of getting a seven given that we have a red card. This yields the result we had before.

The second important thing to note is that conditional probabilities are not always altered by the condition applied. In other words, sometimes the event A makes no difference at all to the probability that B will happen. In such cases

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

This is a form of the 'and' combination of probabilities with which many people are familiar. It is, however, only a special case. Events A and B are such that $P(B \mid A) = P(B)$ are termed *independent* events. For example, suppose we roll a dice several times. The score on each roll should not influence what happens in subsequent throws. If A is the event that I get a 6 on the first roll and B is that I get a 6 on the second, then $P(B)$ is not affected by whether or not A happens, These events are independent. I will discuss some further examples of such events later, but remember for now that independence is a special property and cannot always be assumed.

The final comment I want to make about conditional probabilities is that it does not matter which way round I take the two events A and B. In other words, 'A and B' must be the same as 'B and A'. This means that

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

If we swap the order of my previous logic then we take first the event that my card is a seven. Here $P(B) = 1/13$. Conditioning on this event shrinks the space to only four cards, and the probability of getting a red card in this conditioned space is just $P(A \mid B) = 1/2$. Same answer, different order.

A very nice example of the importance of conditional probability is one that did the rounds in university staff common rooms a few years ago, and recently re-surfaced in Mark Haddon's marvellous novel, *The Curious Incident of the Dog in the Night-Time*. In the version with which I am most familiar it revolves around a very simple game show. The contestant is faced with three doors, behind one of which is a prize. The other two have nothing behind them. The contestant is asked to pick a door and stand in front of it. Then the cheesy host is forced to open one of the other two doors, which has nothing behind it. The contestant is offered the choice of staying where he is or switching to the one remaining door (not the one he first picked, nor the one the host opened). Whichever door he then chooses is opened to reveal the prize (or lack of it). The question is, when offered the choice, should the contestant stay where he is, swap to the other door, or does it not matter?

The vast majority of people I have given this puzzle to answer very quickly that it cannot possibly matter whether you swap or not. But it does. We can see why using conditional probabilities. At the outset you pick a door at random. Given no other information you must have a one-third probability of winning. If you choose not to switch, the probability must still be one-third. That part is easy. Now consider what happens if you happen to pick the wrong door first time. That happens with a probability of two-thirds. Now the host has to show you an empty box, but you are standing in front of one of them so he has to show you the other one. Assuming you picked incorrectly first time, the host has been forced to show you where the prize is: behind the one remaining door. If you switch to this door you will claim the prize, and the only assumption behind this is that you picked incorrectly first time around. This means that your probability of winning using the switch strategy is two-thirds, precisely doubling your chances of winning compared with if you had not switched.

Before we get onto some more concrete applications I need to do one more bit of formalism leading to the most important result in this book, *Bayes' theorem*. In its simplest form, for only two events, this is just a rearrangement of the previous equation

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

The interpretation of this innocuous formula is the seed of a great deal of controversy about the rule of probability in science and philosophy, but I will refrain from diving into the murky waters just yet. For the time being it is enough to note that this is a theorem, so in itself it is not the slightest bit controversial. It is what you do with it that gets some people upset.

This allows you to 'invert' conditional probabilities, going from the conditional probability of A given B to that of B given A. Here is a simple example. Suppose I have two urns, which are indistinguishable from the outside. In one urn (which with a leap of imagination I will call Urn 1) there are 1000 balls, 999 of which are black and one of which is white. In Urn 2 there are 999 white balls and one black one. I pick an urn and am told it is Urn 1. I prepare to draw a ball from it.

I can assign some probabilities, conditional on this knowledge about which urn it is.

Clearly $P(\text{a white ball} \mid \text{Urn 1}) = 1/1000 = 0.001$ and $P(\text{a black ball} \mid \text{Urn 1}) = 999/1000 = 0.999$. If I had picked Urn 2, I would instead assign $P(\text{a white ball} \mid \text{Urn 2}) = 0.999$ and $P(\text{a black ball} \mid \text{Urn 2}) = 0.001$. So far, so good.

Now I am blindfolded and the urns are shuffled about so I no longer know which is which. I dip my hand into one of the urns and pull out a black ball. What can I say about which urn I have drawn from?

Before going on, I have to suppose that some of you will say that I cannot infer anything. I have discussed this problem many times with students and some just seem to be inextricably welded to the idea that you have to have a large number of repeated observations before you can assign a probability. That is not the case. A draw of one ball is enough to say something in this example. Is not it more likely that the ball came from Urn 1 if it is black?

To do this properly using Bayes' theorem is quite easy. What I want is $P(\text{Urn 1} \mid \text{a black ball})$. I have the conditional probabilities the other way round, so it is straightforward to invert them. Let B be the event that I have drawn from Urn 1 and A be the event that the ball is a black one. I want $P(B \mid A)$ and Bayes' theorem gives this as $P(B)P(A \mid B)/P(A)$. I have $P(A \mid B) = 0.999$ from the previous reasoning. Now I need $P(B)$, the probability that I draw a black ball regardless of which urn I picked. The simplest way of doing this is to say that the urns no longer matter: there are just 2000 balls, 1000 of which are white and 1000 of which are black and they are all equally likely to be picked. The probability is therefore $1000/2000 = 1/2$. Likewise for $P(A)$ the balls do not matter and it is just a question of which of two identical urns I pick. This must also be one-half. The required $P(B \mid A) = 0.999$. If I drew a black ball it is overwhelmingly likely that it came from Urn 1.

This gives me an opportunity to illustrate another operation one can do with probabilities: it is called marginalization. Suppose two events, A and B, like before. Clearly B either does or does not happen. This means that when A happens it is either along with B or not along with B. In other words A must be accompanied either by B or by B^*. Accordingly,

$$P(A) = P(A \cap B) + P(A \cap B^*).$$

This can be generalized to any number of mutually exclusive and exhaustive events, but this simplest case makes the point. The first bit, $P(A \cap B) = P(B)P(A \mid B)$, is what appears on the top of the right-hand side of Bayes' theorem, while the second part is just the probability of getting a black ball given that it is not Urn 1. Assuming nobody sneaked any extra urns in while I was not looking this must be Urn 2. The required inverse probability is then $0.999/(0.999 + 0.001)$, as before.

A common situation where conditional probabilities are important is when there is a series of events that are not independent. Card games are rich sources of such examples, but they usually do not involve replacing the cards in the pack and shuffling after each one is dealt. Each card, once dealt, is no longer available for subsequent deals. The space of possibilities shrinks each time a card is removed from the deck, hence the probabilities shift. This brings us to the difficult business of keeping track of the possibility space for hands of cards in games like poker or bridge. This space can be very large, and the calculations are consequently quite difficult.

In the next chapter I discuss how astronomers and physicists were largely responsible for establishing the laws of probability, but I cannot resist the temptation to illustrate the difficulty of combining probabilities by including here an example which is extremely simple, but which defeated the great French mathematician D'Alembert. His question was: in two tosses of a single coin, what is the probability that heads will appear at least once? To do this problem correctly we need to write down the space of possibilities correctly. If we write heads as H and tails as T then there are actually four possible outcomes in the experiment. In order these are HH, HT, TH, and TT. Each of these has the same probability of one-quarter, which one can reckon by saying that each of these pairs must be equally likely if the coin is fair; there are four of them so the probability must be 1/4. Alternatively the probability of H or T is separately 1/2 so each combination has probability 1/2 times 1/2 or 1/4. Three of the outcomes have at least one head (HH, HT, and TH) so the probability we need is just 3/4. This example is very easy because the probabilities in this case are independent, but D'Alembert still managed to mess it up. When he tackled the problem in 1754 he argued that there are in fact only three cases: heads on the first throw, heads on the second throw, or no heads at all. He took these three cases to be equally likely, and deduced the answer to be 2/3. But they are not equally likely: his first case includes two of the correct cases. His possibilities are mutually exclusive, but they are not equally likely.

As an interesting corollary of D'Alembert's error, consider the following problem. A coin is thrown repeatedly in a sequence. Each result is written down. What is the probability that the pair HT appears in the sequence before TT appears? One's immediate reaction to this is to say, like I did before, that HT and TT must be equally likely, so the probability that the one comes before the other must be just 50%. But this is also wrong, because we are not tossing the coin discrete pairs. It is a continuous sequence in which the pairs overlap and are therefore not independent. Suppose my first throw is a head. That has a probability of 50%. Given this starting point, I have to throw the sequence HT before I get TT. If my first throw is a tail then there are two subsequent possibilities: a head next or a tail next. If I through a head next, I have the sequence TH. Again I have to throw a tail to make TT possible down the line somewhere and that inevitably means I have to have THT before I can get, say, THTT. Only if I throw TT right at the start can I ever get TT before HT. The odds are 3 : 1 against this happening.

Now let us get to the serious business of card games, and what they tell us about permutations and combinations. I will start with Poker, because it is the simplest and probably most popular game to lose money on. Imagine I start with a well-shuffled pack of 52 cards. In a game of five-card draw poker, the players bet on who has the best hand made from five cards drawn from the pack. In more complicated versions of poker, such as Texas hold'em, one has, say, two 'private' cards in one's hand and, say, five on the table in plain view. These community cards are usually revealed in stages, allowing a round of betting at each stage. One has to make the best hand one can using five cards from one's private cards and those on the table. The existence of community cards makes this very interesting because it gives some additional information about other player's holdings. For the present discussion, however, I will just stick to individual hands and their probabilities.

How many possible five-card poker hands are there? To answer this question we need to know about permutations and combinations. Imagine constructing a poker hand from a standard deck. The deck is full when you start, which gives you 52 choices for the first card of your hand. Once that is taken you have 51 choices for the second, and so on down to 48 choices for the last card. One might think the answer is therefore $52 \times 51 \times 50 \times 49 \times 48 = 311875200$, but that is not quite the right answer. It does not actually matter in which order your five cards are dealt to you. Suppose you have four aces and the 2 of clubs in your hand. For example, the sequences (A♠, A♦, A♥, A♣, 2♣) and (A♥, A♣, 2♣, A♥, A♦) are counted separately among the number I obtained above. There is quite a large number of ways of rearranging these five cards amongst themselves whilst keeping the same poker hand. In fact, there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ such permutations. Mathematically this kind of thing is denoted 5!, or five-factorial. Dividing the number above by this gives the actual number of possible poker hands: 2,598,960. This number is important because it describes the size of the 'possibility space'. Each of these hands is an elementary outcome of a poker deal, and each is equally likely.

This calculation is an example of a mathematical combination. The number of combinations one can make of r things chosen from a set of n is usually denoted $C_{n,r}$. In the example above, $r = 5$ and $n = 52$. Note that $52 \times 51 \times 50 \times 49 \times 48$ can be written 52!/47! The general result can be written

$$C_{n,r} = \frac{n!}{r!(n-r)!}.$$

Poker hands are characterized by the occurrence of particular events of varying degrees of probability. For example, a 'flush' is five cards of the same suit but not in sequence (2♠, 4♠, 7♠, 9♠, Q♠). A numerical sequence of cards regardless of suit (e.g. 7♥, 8♦, 9♣, 10♥, J♠) is called a straight. A sequence of cards of the same suit is called a straight flush. One can also have a pair of cards of the same value, three of a kind, four of a kind, or a 'full house' which is three of one kind and two of another.

One can also have nothing at all, that is, not even a pair. The relative value of the different hands is determined by how probable they are.

Consider the probability of getting, say, five spades. To do this we have to calculate the number of distinct hands that have this composition. There are 13 spades in the deck to start with, so there are $13 \times 12 \times 11 \times 10 \times 9$ permutations of five spades drawn from the pack, but, because of the possible internal rearrangements, we have to divide again by 5! The result is that there are 1287 possible hands containing five spades. Not all of these are mere flushes, however. Some of them will include sequences too, for example, 8♠, 9♠, 10♠, J♠, Q♠, which makes them straight flush hands. There are only 10 possible straight flushes in spades (starting with 2, 3, 4, 5, 6, 7, 8, 9, 10 or J). So 1277 of the possible hands are flushes. This logic can apply to any of the suits, so in all there are $1277 \times 4 = 5108$ flush hands and $10 \times 4 = 40$ straight flush hands.

I would not go through the details of calculating the probability of the other types of hand, but I have included a table showing their probabilities obtained by dividing the relevant number of possibilities by the total number of hands at the bottom of the middle column. I hope you will be able to reproduce my calculations!

Type of Hand	**Number of Possible Hands**	**Probability**
Straight Flush	40	0.000015
Four of a Kind	624	0.000240
Full House	3744	0.001441
Flush	5108	0.001965
Straight	10,200	0.003925
Three of a Kind	54,912	0.021129
Two Pair	123,552	0.047539
One Pair	1,098,240	0.422569
Nothing	1,302,540	0.501177
Totals	2,598,960	1.00000

Poker involves rounds of betting in which each player tries to assess how likely his hand is to be at the others involved in the game. If your hand is weak, you can fold and allow the accumulated bets to be given to your opponent. Alternatively, you can bluff.

If you bet heavily on your hand, the opponent may well think it is strong even if it contains nothing, and fold even if his hand has a higher value. To bluff successfully requires a good sense of timing—it depends crucially on who gets to bet first—and extremely cool nerves. To spot when an opponent is bluffing requires real psychological insight. These aspects of the game are in many ways more interesting than the basic hand probabilities, and they are difficult to reduce to mathematics.

Another card game that serves as a source for interesting problems in probability is Contract Bridge. This is one of the most difficult card games to play well because it is a game of logic that also involves chance to some degree. Bridge is a game for four people, arranged in two teams of two. The four sit at a table with the two members of each team opposite each other. Traditionally the different positions are called North, South, East, and West, where North and South are partners, as are East and West.

For each hand of Bridge an ordinary pack of cards is shuffled and dealt out by one of the players, the dealer. Let us suppose that the dealer in this case is South. The pack is dealt out one card at a time starting with West (to dealer's left), then North, and so on in a clockwise direction. Each player ends up with 13 cards.

Now comes the first phase of the game, the auction. Each player looks at his cards and makes a bid, which is essentially a coded message that gives information to his partner about how good his hand is. A bid is basically an undertaking to win a certain number of tricks with a certain suit as trumps (or with no trumps). The meaning of tricks and trumps will become clear later. For example, dealer might bid 'one spade' which is a suggestion that perhaps he and his partner could win one more trick than the opposition with spades as the trump suit. This means winning seven tricks, as there are always 13 to be won in a given deal. The next to bid—in this case West—can either pass 'no bid' or bid higher, like an auction. The value of the suits increases in the sequence clubs, diamonds, hearts and spades. So to outbid one spade, West has to bid at least two hearts, say, if hearts is the best suit for him. Next to bid is South's partner, North. If he likes spades as trumps he can raise the original bid. If he likes them a lot he can jump to a much higher contract, such as four spades (4♠). Bidding carries on in a clockwise direction until nobody dares take it higher, Three successive passes will end the auction, and the contract is established. Whichever player opened the bidding in the suit that was chosen for trumps becomes 'declarer'. If we suppose our example ended in 4♠, then it was South that opened the bidding. If West had opened 2♥ and this had passed round the table, West would be declarer.

The scoring system for Bridge encourages teams to go for high contracts rather than low ones, so if one team has the best cards it does not necessarily get an easy ride. It should undertake an ambitious contract rather than stroll through a simple one. In particular there are extra points for making 'game' (a contract of four spades, four hearts, five clubs, five diamonds, or three no trumps). There is a huge bonus available for bidding and making a grand slam (an undertaking to win all thirteen tricks, that is, seven of something) and a smaller but still impressive bonus for a small slam (six of something).

The second phase of the game now starts. The person to the left of declarer plays a card and the player opposite declarer puts all his cards on the table and becomes 'dummy', playing no further part in this particular hand. Dummy's cards are entirely under the control of the declarer. All three players can see them, but only declarer can see his own hand. The card play is then similar to whist. Each trick consists of four cards played in clockwise sequence from whoever leads. Each player, including dummy, must follow the suit led if he has a card of that suit in his hand. If a player does not have a card of that suit he may 'ruff', that is play a trump card, or simply discard something from another suit. One can win a trick in one of two ways. Either one plays a higher card of the same suit, for example, K♥ beats 10♥. Aces are high, by the way. Alternatively one can play a trump. The highest trump played also wins the trick. Note that more than one player may ruff. For instance, East may ruff only to be over-ruffed by South if both have none of the suit led. Of course one may not have any trumps at all, making a ruff impossible. The possibility of winning a trick by a ruff also does not exist if the contract is of the no-trumps variety. Whoever wins a given trick leads to start the next one. This carries on until 13 tricks have been played. Then comes the reckoning of whether the contract has been made. If so, points are awarded to declarer's team. If not, penalty points are awarded to the defenders. Then it is time for another hand, probably another drink, and very possibly an argument about how badly declarer played the hand.

I have gone through the game in some detail in an attempt to make it clear why this is such an interesting game for probabilistic reasoning. During the auction, partial information is given about every player's holding. It is vital to interpret this information correctly if the contract is to be made. The auction can reveal which of the defending team holds important high cards, or whether the trump suit is distributed strangely. Because the cards are played in strict clockwise sequence this matters a lot. On the other hand, even with very firm knowledge about where the important cards lie, one still often has a difficult logical puzzle to solve if all of one's winners are to be made. It can be a very subtle game.

I have only space here for one illustration of this kind of thing, but it is one that is fun to work out. As is true to a lesser extent in poker, one is not really interested in the initial probabilities of the different hands but rather how to update these probabilities using conditional information as it may be revealed through the auction and card play. In poker this updating is done largely by interpreting the bets one's opponents are making.

Let us suppose that I am South, and I have been daring enough to bid a grand slam in spades (7♠). West leads, and North lays down dummy. I look at my hand and dummy, and realize that we have 11 trumps between us, missing only the King and the 2. I have all other suits covered, and enough winners to make the contract provided I can make sure I win all the trump tricks. The King, however, poses a problem. The Ace of Spades will beat the King, but if I just lead the Ace, it may be that one of East or West has both the K♠ and the 2♠. In this case he would simply play the two to my Ace. The King would be an automatic winner then: as the highest remaining trump it must win a trick eventually. The contract is then lost. Of course if the spades are split 1-1 between East and West then the King drops when I lead the Ace, so that works.

But there is a different way to play this situation. Suppose, for example, that A♠ and Q♠ are on the table and I have managed to win the first trick in my hand. If I think the K♠ lies in West's hand, I lead a spade. West has to play a spade. If he has the King, and plays it, I can cover it with the Ace so it does not win. If, however, West plays low I can play Q♠. This will win if I am right about the location of the King. Next time I can lead the A♠ from dummy and the King falls. This play is called a *finesse*. But is this better than playing for the drop? It is all a question of probabilities, and this in turn boils down to the number of possible deals that allow each strategy to work.

To start with, we need the total number of possible bridge hands. This is quite easy: it is the number of combinations of 13 objects taken from 52, that is $C_{52,13}$. This is a truly enormous number: over 600 billion. You have to play a lot of games to expect to be dealt the same hand twice!

What we now have to do is evaluate the probability of each possible arrangement of the missing King and two. Dummy and declarer's hands are known to me. There are 26 remaining cards whose location I do not know. The relevant space of possibilities is now smaller than the original one. I have 26 cards to assign between East and West. There are $C_{26,13}$ ways of assigning West's 13 cards, but once I have done this the remaining 13 must be in East's hand.

Suppose West has the 2 but not the K. Conditional on this assumption, I know one of his cards, but there are 12 others remaining to be assigned. There are therefore $C_{24,12}$ hands with this possible arrangement of the trumps. Obviously the K has to be with East in this case. The opposite situation, with West having the K but not the 2 has the same number of possibilities associated with it. Suppose instead West does not have any trumps. There are $C_{24,13}$ ways of constructing such a hand: 13 cards from the 24 remaining non-trumps. The remaining possibility is that West has both trumps: this can happen in $C_{24,11}$ ways. To turn these counts into probabilities we just divide by the total number of different ways I can construct the hands of East and West, which is $C_{26,13}$.

Spades in West's Hand	**Number of Hands**	**Probability**	**Drop**	**Finesse**
None	$C_{24,13}$	0.24	0	0
K	$C_{24,12}$	0.26	0.26	0.26
2	$C_{24,12}$	0.26	0.26	0
K2	$C_{24,11}$	0.24	0	0.24
Total	$C_{26,13}$	1.00	0.52	0.50

The last two columns show the contributions of each arrangement to the probability of success of either playing for the drop or the finesse. The drop is slightly more likely to work than the finesse in this case. Note, however, that this ignores any information gleaned from the auction, which could be crucial. Note also that the probability of the drop and the probability of the finesse do not add up to one. This is because there are situations where both could work or both could fail.

This calculation does not mean that the finesse is never the right tactic. It sometimes has much higher probability than the drop, and is often strongly motivated by information the auction has revealed.

Calculating the odds precisely, however, gets more complicated the more cards are missing from declarer's holding. For those of you too lazy to compute the probabilities, the book *On Gambling*, by Oswald Jacoby contains tables of the odds for just about any bridge situation you can think of.

Finally on the subject of Bridge, I wanted to mention a fact that many people think is paradoxical but which is really just a more complicated version of the 'three-door' problem I discussed above. Looking at the table shows that the odds of a 1-1 split in spades here are 0.52 : 0.48 or 13 : 12. This comes from how many cards are in East and West's hands when the play is attempted. There is a much quicker way of getting this answer than the brute force method I used above. Consider the hand with the spade 2 in it. There are 12 remaining opportunities in that hand that the spade K might fill, but there are 13 available slots for it in the other. The odds on a 1-1 split must therefore be 13 : 12. Now suppose instead of going straight for the trumps, I play off a few winners in the side suits (risking that they might be ruffed, of course). Suppose I lead out three Aces in the three suits other than spades and they all win. Now East and West have only 20 cards between them and by exactly the same reasoning as before, the odds of a 1-1 split have become 10 : 9 instead of 13 : 12. Playing out seemingly irrelevant suits has increased the probability of the drop working. Although I have not touched the spades, my assessment of the probability has changed significantly.

I want to end this Chapter with a brief discussion of some more mathematical (as opposed to arithmetical) aspects of probability. I will do this as painlessly as possible using two well-known examples to illustrate the idea of probability distributions and random variables. This requires mathematics that some readers may be unfamiliar with, but it does make some of the examples I use later in the book a little easier to understand.

In the examples I have discussed so far I have applied the idea of probability to discrete events, like the toss of a coin or a ball drawn from an urn. In many problems in statistical science the event boils down to a measurement of something, that is, the numerical value of some variable or other. It might be the temperature at a weather station, the speed of a gas molecule, or the height of a randomly-selected individual. Whatever it is, let us call it X. What one needs for such situations is a formula that supplies the relative probability of the different values X can take. For a start let us assume that X is *discrete*, that is, that it can only take on specific values. A common example is a variable corresponding to a count (the score on a dice, the number of radioactive decays recorded in a second, and so on). In such cases X is an integer, and the possibility space is $\{0, 1, 2, \ldots\}$. In the case of a dice the set is finite $\{1, 2, 3, 4, 5, 6\}$ while in other examples it can be the entire set of integers going up to infinity.

The probability distribution, $p(x)$, gives the probability assigned to each value of X. If I write $P(X=x)=p(x)$ it probably looks unnecessarily complicated, but this means that 'the probability of the random variable X taking on the particular numerical value x is given by the mathematical function $p(x)$'. In cases like this we use the probability laws in a slightly different form. First, the sum over all probabilities must be unity:

$$\sum_x p(x) = 1,$$

If there is such a distribution we can also define the expectation value of X, $E(X)$ using

$$E(X) = \sum_x xp(x)$$

The expectation value of any function of X, say $f(X)$, can be obtained by replacing x by $f(x)$ in this formula so that, for example:

$$E(X^2) = \sum_x x^2 p(x).$$

A useful measure of the spread of a distribution is the variance, usually expressed as the square of the standard deviation, σ, as in

$$\sigma^2(X) = E(X^2) - [E(X)]^2.$$

To give a trivial example, consider the probability distribution for the score X obtained on a roll of a dice. Each score has the same probability, so $p(x)=1/6$ whatever x is. The formula for the expectation value gives

$$\begin{aligned} E(X) &= 1 \times 1/6 + 2 \times 1/6 + 3 \times 1/6 + 4 \times 1/6 + 5 \times 1/6 \\ &\quad + 6 \times 1/6 \\ &= 21/6 = 3.5 \end{aligned}$$

Incidentally, I have never really understood why this is called the expectation value of X. You cannot expect to throw 3.5 on a dice—it is impossible! However, it is what is more commonly known as the average, or arithmetic mean. We can also see that

$$\begin{aligned} E(X^2) &= 1 \times 1/6 + 2^2 \times 1/6 + 3^2 \times 1/6 + 4^2 \times 1/6 + 5^2 \times 1/6 \\ &\quad + 6^2 \times 1/6 \\ &= 91/6 \end{aligned}$$

This gives the variance as $91/6 - (21/6)^2$, which is 35/12. The standard deviation works out to be about 1.7. This is a useful thing as it gives a rough measure of the spread of the distribution around the mean. As a rule of thumb, most of the probability lies within about two standard deviations either side of the mean.

Let us consider a better example, and one which is important in a very large range of contexts. It is called the binomial distribution. The situation where it is relevant is when we have a sequence of n independent 'trials' each of which has only two possible outcomes ('success' or 'failure') and a constant probability of 'success' p. Trials like this are usually called Bernoulli trials, after Daniel Bernoulli who is discussed in the next chapter. We ask the question: what is the probability of exactly x successes from the possible n? The answer is the binomial distribution:

$$p_n(x) = C_{n,x} p^x (1-p)^{n-x}$$

You can probably see how this arises. The probability of x consecutive successes is p multiplied by itself x times, or p^x. The probability of $(n - x)$ successive failures is $(1 - p)^{n-x}$. The last two terms basically therefore tell us the probability that we have exactly x successes (since there must be $n - x$ failures). The combinatorial factor in front takes account of the fact that the ordering of successes and failures does not matter. For small numbers n and x, there is a beautiful way called Pascal's triangle, to construct the combinatorial factors. It is cumbersome to use this for large numbers, but in any case these days one can use a calculator.

The binomial distribution applies, for example, to repeated tosses of a coin, in which case p is taken to be 0.5 for a fair coin. A biased coin might have a different value of p, but as long as the tosses are independent the formula still applies. The binomial distribution also applies to problems involving drawing balls from urns: it works exactly if the balls are replaced in the urn after each draw, but it also applies approximately without replacement, as long as the number of draws is much smaller than the number of balls in the urn. It is a bit tricky to calculate the expectation value of the binomial distribution, but the result is not surprising: $E(X) = np$. If you toss a fair coin 10 times the expectation value for the number of heads is 10 times 0.5, which is 5. No surprise there. After another bit of maths, the variance of the distribution can also be found. It is $np(1 - p)$.

The binomial distribution drives me insane every four years or so, whenever it is used in opinion polls. Polling organisations generally interview around 1000 individuals drawn from the UK electorate. Let us suppose that there are only two political parties: Labour and the rest. Since the sample is small the conditions of the binomial distribution apply fairly well. Suppose the fraction of the electorate voting Labour is 40%, then the expected number of Labour voters in our sample is 400. But the variance is $np(1 - p) = 240$. The standard deviation is the square root of this, and is consequently about 15. This means that the likely range of results is about 3% either side of the mean value. The 'term' 'margin of error' is usually used to describe this sampling uncertainty. What it means is that, even if political opinion in the population at large does not change at all the results of a poll of this size can differ by 3% from sample to sample. Of course this does not stop the media from making stupid statements like 'Labour's lead has fallen by 2%'. If the variation is within the margin of error then there is absolutely no evidence that the proportion p has changed at all. Doh!

So far I have only discussed discrete variables. In the physical sciences one is more likely to be dealing with continuous quantities, that is, those where the variable can take any numerical value. Here we have to use a bit of calculus to get the right description: basically, instead of sums we have to use integrals. For a continuous variable, the probability is not located at specific values but is smeared out over the whole possibility space. We therefore use the term probability density to describe this situation. The probability density $p(x)$ is such that the probability that the random variable X takes a value in the range $(x, x + dx)$ is $p(x)\,dx$. The density $p(x)$ is therefore not a probability itself, but a probability *per unit* x. With this definition we can write

$$\int_x p(x)\, dx = 1.$$

The probability that X lies in a certain range, say $[a, b]$, the area under the curve defined by $p(x)$:

$$P(a \leq x \leq b) = \int_a^b p(x)\, dx.$$

Expectation values are defined in an analogous way to the case of discrete variables, but replacing sums with integrals. For example,

$$E(X) = \int_x xp(x)\, dx.$$

I have really included these definitions for completeness. Do not worry too much if you do not know about differential calculus, as I will not be doing anything difficult along these lines. This formalism does however allow me to introduce what is probably the most important distribution in all probability theory. This is the Gaussian distribution, often called the normal distribution. It plays an important role in a whole range of scientific settings. This distribution is described by two parameters: μ and σ, of which more in a moment. The mathematical form is

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$

but it is only really important to recognize the shape, which is the famous 'Bell Curve' shown in the Figure. The expectation value of X is $E[X] = \mu$ and the variance is σ^2.

So why is the Gaussian distribution so important? The answer is found in a beautiful mathematical result called the *Central Limit Theorem*. This used to be called the 'Law of Frequency of Error', but since it applies to many more useful things than errors I prefer the more modern name, This says, roughly speaking, that if you have a variable, X, which arises from the sum of a large number of

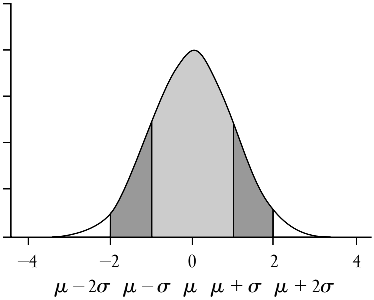

Figure 2 The Normal distribution. The peak of the distribution is at the mean value (μ), with about 95% of the probability within 2σ on either side

independent random influences, so that

$$X = X_1 + X_2 + \cdots X_n$$

then whatever the probabilities of each of the separate influences X_i, the distribution of X will be close to the Gaussian form. All that is required is that the X_i should be independent and there should be a large number of them. Note also that the distribution of the sum of a large number of a independent Gaussian variables is exactly Gaussian. There are an enormous number of situations in the physical and life sciences where some effect is the outcome of a large number of independent causes. Heights of individuals drawn from a population tend to be normally distributed. So do measurement errors in all kinds of experiments. In fact, even the distribution resulting from a very large number of Bernoulli trials tends to this form. In other words, the limiting form of the binomial distribution for a very large n is itself of the Gaussian form, with μ replaced by np and σ^2 replaced by $np(1-p)$. This does not mean that everything is Gaussian. There are certainly many situations where the central limit theorem does not apply, but the normal distribution is of fundamental importance across all the sciences. The Central Limit Theorem is also one of the most remarkable things in modern mathematics, showing as it does that the less one knows about the individual causes, the surer one can be of some aspects of the result. I cannot put it any better than Sir Francis Galton:

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed

> by the 'Law of Frequency of Error'. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.

References and Further Reading

For a good introduction to probability theory, as well as its use in gambling, see:

Haigh, John. (2002). *Taking Chances: Winning with Probability*, Second Edition, Oxford University Press.

A slightly more technical treatment of similar material is:

Packel, Edward. (1981). *The Mathematics of Games and Gambling*, New Mathematical Library (Mathematical Association of America).

More technically mathematical works for the advanced reader include:

Feller, William. (1968). *An Introduction to Probability Theory and Its Applications*, Third Edition, John Wiley & Sons.

Grimmett, G.R. and Stirzaker, D.R. (1992). *Probability and Random Processes*, Oxford University Press.

Jaynes, Ed. (2003). *Probability Theory: The Logic of Science*, Cambridge University Press.

Jeffreys, Sir Harold. (1966). *Theory of Probability*, Third Edition, Oxford University Press.

Simple applications of probability to statistical analysis can be found in

Rowntree, Derek. (1981). *Statistics without Tears*, Pelican Books.

Finally, you must read the funniest book on statistics, once reviewed as 'wildly funny, outrageous, and a splendid piece of blasphemy against the preposterous religion of our time':

Huff, Darrell. (1954). *How to Lie with Statistics*, Penguin Books.