# CHAPTER 4

Probability

While the graphical and numerical methods of Chapters 2 and 3 provide us with tools for summarizing data, probability theory, the subject of this chapter, provides a foundation for developing statistical theory. Most people have an intuitive feeling for probability, but care is needed as intuition can lead you astray if it does not have a sure foundation.

After talking about the meaning of events and probability we introduce a number of rules for calculating probabilities. Important ideas that we develop are conditional probability and the concept of independent events. We will exploit two aids in developing these ideas. The first is the two-way table of counts or proportions which enables us to understand in a simple way how many probabilities encountered in real-life are computed. The second is the tree diagram. This is often useful in clarifying our thinking about sequences of events. A number of case studies will highlight some practical features of probability and its possible pitfalls.

4.1   Introduction

"If I toss a coin, what is the probability that it will turn up heads?" It's a rather silly question isn't it? Everyone knows the answer, namely "a half" or "one chance in two" or "fifty-fifty". But let us look a bit more deeply behind this response that everyone makes.
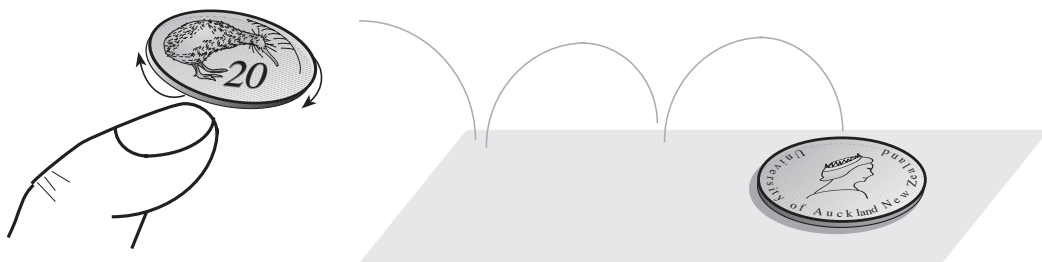
Firstly, why is the probability one half? When you ask a class, a dialogue like the following often develops. "The probability is one half because the coin is equally likely to come down heads or tails." Well, it could conceivably land on its edge but we can fix that by tossing again. So why are the two outcomes "heads" and "tails" equally likely? "Because it's a fair coin." Sounds like somebody had taken statistics before, but that's just jargon isn't it. What does it mean? "Well it's symmetrical." Have you ever seen a symmetrical coin? They are always different on both sides with different bumps and indentations. These might influence the chances that the coin comes down heads. How could we investigate this? "We could toss the coin lots of times and see what happens." This leads us to an intuitively attractive approach to probabilities for repeatable "experiments" such as coin tossing or die rolling: probabilities are described in terms of long run relative frequencies from repeated trials. Assuming these relative frequencies become stable after a large enough number of trials, the probability could be defined as the limiting relative frequency. Well it turns out that several people have tried this.
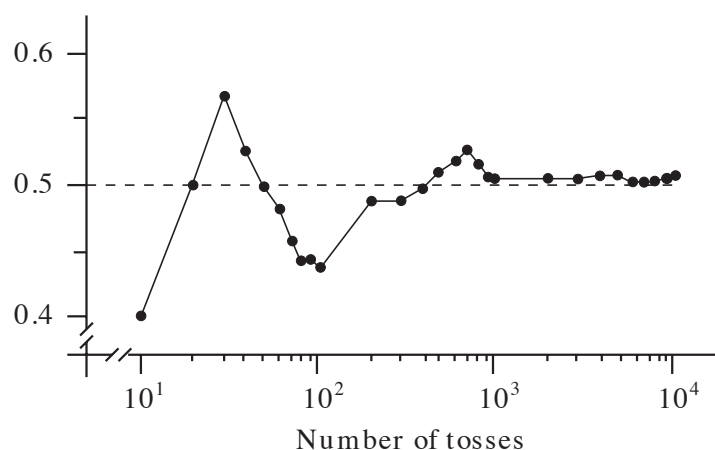

Figure 4.1.1 : Proportion of heads versus number of tosses for John Kerrich's coin tossing experiment.

English mathematician John Kerrich was lecturing at the University of Copenhagen when World War II broke out. He was arrested by the Germans and spent the war interned in a camp in Jutland. To help pass the time he performed some experiments in probability. One of these involved tossing a coin ten thousand times and recording the results.[1] Fig. 4.1.1 graphs the proportion of heads Kerrich obtained up to and including toss numbers $10, 20, \ldots, 100$, $200, \ldots, 900, 1,000, 2,000, \ldots, 9,000, 10,000$. The data is given in Freedman et al. [1991, Table 1, p. 248].

For the coin Kerrich was using, the proportion of heads certainly seems to be settling close to $\frac{1}{2}$. This is an *empirical* or experimental approach to probability.

---

[1] John Kerrich hasn't been the only person with time on his hands. In the eighteenth century, French naturalist Comte de Buffon performed $4,040$ tosses for $2,048$ heads while around 1900, Karl Pearson one of the pioneers of modern statistics, made $24,000$ tosses for $12,012$ heads.

We never know the exact probability this way, but we can get a pretty good estimate. Kerrich's data gives us an example of the frequently observed fact that in a long sequence of independent repetitions of a random phenomenon, the proportion of times in which an outcome occurs gets closer and closer to a fixed number which we can call the probability that the outcome occurs. This is the *relative frequency* definition of probability. The long run predictability of relative frequencies follows from a mathematical result called the Law of Large Numbers.

But let's not abandon the first approach via symmetry quite so soon. The answer it gave us agreed well with Kerrich's experiment. It was based upon an artificial idealization of reality – what we have called a *model*. The coin is imagined as being completely symmetrical so that there should be no preference for landing heads or tails. Each should occur half the time. However, no coin is exactly symmetrical nor, in all likelihood, has a probability of landing heads of exactly $\frac{1}{2}$. Yet experience has told us that the answer the model gives is close enough for all practical purposes. Although real coins are not symmetrical, they are close enough to being symmetrical for the model to work well.

4.2 Coin Tossing and Probability Models

At a University of London seminar series, NZ statistician Brian Dawkins asked the very question we used to open this discussion. He then left the stage and came back with an enormous coin, almost as big as himself.

The best he could manage was a single half turn. There was no element of chance at all. Some people can even make the tossing of an ordinary coin predictable. Program 15 of the *Against All Odds* video series (COMAP[1989]) shows Harvard probabilist Persi Diaconis tossing a fair coin so that it lands heads every time.[2] It is largely a matter of being able to repeat the same action. By just trying to make the action very similar we could probably shift the chances from 50 : 50 but most of us don't have that degree of interest or muscular control. A model in which coins turn up heads half the time and tails the other half in a totally unpredictable order is an excellent description of what we see when we toss a coin. Our physical model of a symmetrical coin gives rise to a *probability model* for the experiment. A probability model has two essential components: the *sample space*, which is simply a list of all

---

[2]Diaconis was a performing magician before training as a mathematician.

the outcomes the experiment can have, and a list of *probabilities*, where each "probability" is intended to be the probability that the corresponding outcome occurs (see Section 4.4.4).

In sport, coins are tossed to decide which end of the ground a team is to defend, or who is going to go into bat first. It is quite common to call the outcome after the coin has been tossed, but before it has fallen. Even with a normal coin, the side that it lands on is virtually completely determined by a number of factors such as which way up it started, the degree of spin, the speed and angle with which it left the thumb and how far it has to fall. If we knew all this, then, with sufficient expertise in physics, we could write down some equations which are thought to govern the motion of the coin (a mathematical model like Newton's laws) and use these to work out which way up the coin should land. The mathematical model will give us precise predictions but it has some drawbacks. It will be complicated and it will require us to have some way of accurately measuring quantities such as speeds and angles. Our probability model, on the other hand, is simple and requires no such information. But we pay for this by not being able to predict specific occurrences (e.g. whether this toss will result in a head). We also need to assume that the various physical factors, like those mentioned above, vary in an unpredictable fashion.

Probability models can only talk about average behavior in the long run. In a very long sequence of coin tosses we will see very nearly the same *proportions* (or relative frequencies) of heads and tails. This is an example of what is popularly referred to as the "law of averages"[3]. Many misconceptions have developed around the idea of the "law of averages". Even though the proportion (relative frequency) of heads becomes more and more stable as the number of tosses increases, inspection of Kerrich's data in Freedman et al. [1991, p. 248] reveals that the difference between the actual numbers of heads and tails becomes more and more variable. The law of averages applies to relative frequencies, not to absolute numbers. Many people believe that after a sequence of 10 straight heads the chances of getting a tail is much bigger than 50% "because the law of averages will be trying to even things up." The law of averages says nothing at all about short run behavior. If you inspect very long sequences of coin tosses, find all instances of 10 heads in a row and then look at what happens next each time, you will find that the next toss is still a tail about half of the time and a head about half of the time. Short term behavior is utterly unpredictable. And, of course, these are precisely the reasons coins are used to start sports games – the results are "fair" but unpredictable.[4]

Let us return to our mathematical model for coin tossing. Even if we could make all the measurements required by a mathematical model of coin tossing, it is unlikely that it would always give us the correct answers. There would be factors affecting the experiment that we had not allowed for. For example, a

---

[3]This "law" is the popular version of the technical *Law of Large Numbers* discussed in the alternate version of Chapter 5 given on the web site.

[4]For a more detailed exposition of these ideas, see Freedman et al. [1991, Chapter 6] and Moore [1991, pages 331-339].

gust of wind or slight variations in the degree of bounce in the table could still render our answer wrong some of the time. Furthermore, measurements can never be made completely accurately. Measurement errors are often random in appearance and may be large enough to adversely affect the answers. In building probability models for real world situations we need to model both the predictable factors that we know about using mathematics, and the unpredictable factors, using probability. Some of the unpredictable factors may or may not actually be random. This does not matter. The important thing is that they are unpredictable to us and look random. We therefore describe these unpredictable elements in terms of random events with associated probabilities. A model which consists of a predictable ("deterministic") part and an unpredictable ("stochastic") part is typically used when we fit a straight line to data. We model the predictable part, the pattern or trend, by a straight line, and we model the unpredictable part, namely the variation of the points about the line, using ideas of randomness and probability. This more complex type of model is discussed later in Chapter 12. In the meantime, however, we shall confine ourselves to very simple situations.

Another two-outcome probability model

The gender of a child is a two-outcome "experiment" whose outcome still appears to be random. With current scientific knowledge it is still impossible to *control* the sex of the conceived child. This is despite Aristotle's belief that boys tend to be produced if the father is highly excited and other venerable solutions such as waiting until the wind is in the north, keeping one's boots on, and eating raw eggs. To the best of our knowledge, none of them work reliably! Conception can be viewed as a race. The winner is the first sperm to reach the egg and still have enough energy to penetrate its wall. Does it carry an X chromosome resulting in a girl, or carry a Y, giving a boy? Looking around, and seeing roughly equal numbers of males and females we may decide that it is just like tossing a coin. Thus we could use the same probability model, but with boys and girls replacing heads and tails. The set of possible outcomes is {boy, girl} and each outcome has a probability of a $\frac{1}{2}$. This model works reasonably well but it has deficiencies. In most countries the birth rate for boys is slightly higher than that for girls. In NZ, for example, roughly 52% of births are boys.

Another deficiency in the analogy between tossing coins and having children becomes apparent when we come to have the second child. The chances of getting a girl should be the same whether or not the first child was a girl (after all, the coin doesn't know whether it came down heads or tails last time). This idea is called independence. However there is some evidence that people whose first child is a girl or a boy have a second child of the same sex slightly more frequently than one of the opposite sex. Having made these points, however, we note that the deficiencies of the coin tossing model as a description of child's gender order in families are slight and that probabilities from the coin tossing model are accurate enough for almost all practical purposes.

1. What does the law of averages say about the behavior of coin tosses?
2. What are two widely held misconceptions about what the law of averages says about coin tosses?
3. Describe two ways in which the coin tossing model is inadequate for describing the gender order of children in families. Are the deficiencies big enough to be important?

4.3 Where do Probabilities come from?

In the previous section, we stated that a probability model had two essential components, the sample space which is a list of all the possible outcomes our experiment can have, and a list of probabilities, one for each outcome. But where do the probabilities that we meet in everyday life come from? We have seen some examples in the previous section. Here is another one.

In 1977 a PanAm jumbo jet and a *KLM* jumbo jet collided on an airport runway in the Canary Islands. One jet was taxiing after landing while the other was taking off. Five hundred and eighty one lives were lost. Soon after, well-known Australian statistician, Terry Speed, noticed the following wire service report in *The West Australian*.

> "NEW YORK, Mon: Mr. Webster Todd, Chairman of the American National Transportation Safety Board said today statistics showed that the chances of two jumbo jets colliding on the ground were about 6 million to one .....-AAP."

Many people are frightened of flying and major air disasters increase such fears. It seems clear from the report that the National Transportation Safety Board has responded with a scientifically based assessment based upon hard data ("statistics showed ...") of the chances of such an accident occurring again so that people could put their fears into perspective. Terry Speed, who has strong research interests in probability, was intrigued by this and wondered how the Board had calculated their figure. So Terry wrote to the Chairman. He received the following reply from a high government official which we reproduce with his permission:

> Dear Professor Speed,
>
>   In response to your aerogram of April 5, 1977, the Chairman's statement concerning the chances of two jumbo jets colliding (6 million to one) has no statistical validity nor was it intended to be a rigorous or precise probability statement. The statement was made to emphasize the intuitive feeling that such an occurrence indeed has a very remote but not impossible chance of happening.

Thank you for your interest in this regard.

Sincerely yours, etc.

At best, the quoted probability was a subjective assessment. At worst, it was a vanishingly small number plucked out of thin air to reassure the public.[5]

We have seen examples now of each of the three main ways that probabilities are assigned to events: (a) from models, (b) from data, and (c) subjectively. These different ways are now described in detail.

4.3.1 Probabilities from models

We can sometimes think up a sufficiently simple model of a real experiment in which it is easy to determine a probability. The simplest cases of this occur when the model leads us to believe that the outcomes are equally likely. This is why we believe that the probability of getting a head when tossing a coin is $\frac{1}{2}$, that the chances of any particular outcome (say a 4) on rolling a standard die is $\frac{1}{6}$, and that the chances of drawing any particular card (say ace of hearts) from a standard deck of cards is 1 in 52. Unfortunately, the method tends to be limited to a few special cases which are simple enough to be treated in this way. Now we know that the probabilities that the model gives will only be approximately true for the real experiment. If the assumptions of the model are sufficiently wrong, the answers can be completely wrong. Let us think about card games. The probabilities given for card games depend critically upon the cards being "well shuffled" so that their order is "random". From experience we know that children and beginners can't shuffle cards very well. Some clumps of cards don't get mixed up much. Experts are so clever at shuffling that is hard for a layperson to know whether they are shuffling for randomness or shuffling to their own advantage. There are also many famous examples of where supposedly random lottery draws have exhibited behavior which is clearly nonrandom. Perhaps the most famous example is the United States military draft lottery of 1970 during the Vietnam War.

---

[5]The quoting of vanishingly small probabilities for air disasters is certainly not a thing of the past. In 1989, a brand new British Midlands twin-engine Boeing 737–400 airliner crashed near Kegworth in Central England killing 44 people. Early reports stated that both engines failed. The *NZ Herald* (11 January 1989, page 6) quoted Captain John Guntrip, an official of the British Guild of Air Pilots as saying "The odds of this happening are astronomical – about one chance in one hundred million."
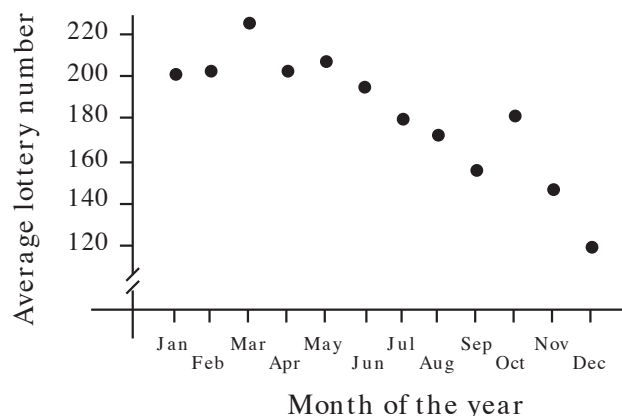
Figure 4.3.1 : Average lottery numbers by month.
Replotted from data in Fienberg [1971].

The military draft was based upon the dates of birth of eighteen-year-old men and worked as follows. Three hundred and sixty six identical cylinders were used, each containing a slip of paper with a day of the year written upon it (1952, the men's birth year, was a leap year). The cylinders were poured into a two foot bowl, supposedly randomly mixed. An official drew them out one by one. The order of drawing was the *draft number*. Thus if June 10 was drawn 5th, all eighteen-year-old males born on June 10 had the draft number 5. The actual draft was performed by conscripting all of these men who had a draft number lower than some limit where the limit was set to fulfill the quota for soldiers for that year. Recall that most of these soldiers would end up fighting in jungles of Vietnam so that randomness of draw was important in the interests of fairness. Everyone should have the same chance of being drafted. However, in 1970, reporters noted that men born later in the year tended to have lower draft numbers (and therefore a greater chance of being drafted) than those born earlier in the year (see Fig. 4.3.1). Statistical calculations revealed that such a strong association between draft number and birthdate could be expected to occur less than once every thousand years with truly random lotteries. Moreover, from a close inspection of the mixing process which was devised by a Colonel Fox and a Captain Pascoe, one might have expected that December cylinders would tend to be closer to the top of the bowl than those for January, say.

The job of devising the lottery was subsequently taken from the military and given to statisticians at the US National Bureau of Standards who devised a scheme based on various levels of randomization. Dates were placed into date capsules in random order determined by a table of random numbers. The date-cylinders were placed in a drum in random order (using another table). Draft numbers (1 to 365 unless a leap year) were placed in a second drum in random order (using a third table). Both drums were rotated for an hour. In front of T.V. cameras, an official then pulled out cylinders in pairs, one from each drum being the date of birth and its associated draft number. As Moore

[1991] says, "it's awful, but it's random". For further details see Moore [1991 : pages 62–65] or Fienberg [1971], and Rosenblatt and Filliben [1971].

4.3.2   Probabilities from data

If we did not have a model for coin tossing that was so obviously a good one, a natural approximation for the probability of John Kerrich's coin coming up heads would be the proportion Kerrich observed, namely $5067/10,000$. Moreover, most people would transfer this figure to their own coins because they could see no reason why their coin should behave differently from Kerrich's. Because the probability of an outcome is the long run relative frequency if we can independently repeat the experiment over and over again, the bigger the sample observed the more reliable the answer. We would have more faith in Kerrich's figure based upon $10,000$ tosses than a figure based upon only 100 or even $1,000$ tosses.[6]

A major source of quoted probabilities for events is data on the relative frequencies of these same events in the past. In New Zealand, last year roughly 700 people from a population of about 3 million were killed on the roads. Most people would be fairly comfortable with a statement of the form "the probability that a randomly selected New Zealander will die on the roads next year" is about $700/3,000,000$. There are two important considerations however. Firstly, we can only make such statements if we think that the underlying process is stable over time. For example, if we knew that the open road speed limit would be increased before the end of next year, our estimate would have to be revised upwards and it is by no means clear how we should do this. Secondly, as noted above, our relative frequencies have to be taken from large numbers for us to have much confidence in them as probabilities. A real success story of the use of historical relative frequencies to provide probabilities of similar events in the future is provided by the life insurance industry. Life companies need good estimates of the chances that various types of people will die within a given period so that they can calculate what premiums have to be charged to cover the maximum probable levels of claims. Referring back to the wire services report of Section 4.3, the US National Transportation Safety Board could conceivably have quoted a relative frequency probability. Every day, hundreds of commercial airliners take off and land all around the world. The Safety Board could have estimated the number of takeoffs and landings in the previous 10 years say, counted the number of runway collisions and quoted a figure like 1 runway collision per 6 million takeoffs. And many newspaper readers will probably have interpreted the Board's (purely subjective) statement in such terms.

There are two important relationships between probabilities from theoretical models and probabilities based upon relative frequency data. Firstly, if the model is reasonable, then for any experiment that can be repeated over and

---

[6]We will see how the estimate of a probability taken from a relative frequency becomes more accurate as the number of repetitions increases in Chapter 7.

over again, the probability of an event obtained from the model tells us the relative frequency with which that event will occur over the long run. Secondly, we may choose not to use the observed relative frequencies as our probabilities, but simply use them to make us feel happier about our probabilities obtained from the model. (Either that or tell us to throw away the model!) For example, suppose we observed 24 heads in 40 tosses of some particular coin. Few people would then use $\frac{24}{40}$ for the probability of getting a head for that coin. Most would observe that 24 in 40 was reasonably in line with the model-based value of $\frac{1}{2}$ and thus feel happier about using this value in future.

4.3.3 Subjective probabilities

In 1992, Lil E. Tee won the Kentucky Derby. Suppose that an ordinary racegoer had thought that the chances of Lil E. Tee winning the Derby were $\frac{3}{4}$. What would he or she mean? It doesn't make sense to think of it in terms of the 1992 Kentucky Derby being run many times and Lil E. Tee winning three quarters of those races. The punter doesn't make this assessment on the basis that Lil E. Tee has won three quarters of its past races. The serious punter will make an assessment from a subjective pooling of all relevant information that he or she may be aware of including Lil E. Tee's past record, the records of other horses in the race, the state of the track and any information picked up about the current form of the horses. But basically it comes down to a numerical measure of the strength of that punter's belief in the proposition that Lil E. Tee will win. Another punter exposed to the same sources of information would have a different strength of belief. Some punters may have a strong sense of belief and thus give the proposition a high probability for what appear, to most of us, to be ludicrous reasons.

In contrast to subjective probabilities, most people can agree about relative frequency probabilities and even about model probabilities if they are backed up by data. Unfortunately, subjective probabilities often masquerade as frequency probabilities. Table 4.3.1, taken from Speed [1977], was abridged from a table in the Reactor Safety Study (or Rasmussen Report), a major US governmental study of the safety of nuclear power facilities.

Table 4.3.1 : Fatality Rates[a]

Accident	Average Chance of Death per Year per individual	
Motor Travel	1 in	3,000
Air Travel	1 in	100,000
Reactor accidents (based on 100 reactors in the US)		
(a)　within a few weeks	1 in	300,000,000
(b)　within about twenty weeks	1 in	16,000,000

[a]Source: Speed [1977]. Abridged from a table in the Reactor Safety Study cited by Speed.

The first two probabilities are frequency probabilities based upon plenty of hard data. One out of every 3,000 of the millions of people in the US die on the roads per year. One in every 100,000 dies in an air accident in a year. The juxtaposition with these figures makes it appear that the following reactor accident values are similarly frequency probabilities based upon data. In fact, they are based upon calculations of the likelihoods of chains of events (see Section 4.7.3). Many of the individual probabilities in the chain were somebody's subjective assessment so that the result is really only a subjective probability.

4.3.4 Manipulation of probabilities

While not all statisticians will agree about what probabilities should be associated with a particular real world event, all do agree on how probabilities should be combined and manipulated. This is the subject of the following section. It even applies to subjective probabilities. Subjective Bayesian statisticians, who believe that the process of statistical inference should be concerned with using data to refine one's subjective degree of belief in a theory or statement, still use the ordinary rules of probability for this refinement process.

Quiz on Section 4.3

1. What are the three types of probability we typically encounter?
2. Give examples of each from your own experience.
3. What assumption underlies probabilities given for card games?
4. When the relative frequency of an event in the past is used to estimate the probability it occurs in the future, what assumption is being made?
5. What do all statisticians agree about with respect to probabilities (Section 4.3.4)?
6. When a weather forecaster says that there is a 70% chance of rain tomorrow, what do you think this statement means?

Exercises on Section 4.3

1.   Suppose we make a spinner as shown in the picture. The experiment is to spin the pointer vigorously and see what color it stops on. How would you obtain a relative frequency probability for the probability that it stops on grey?
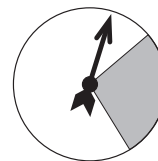
Can you calculate a model-based probability of stopping on grey? If so how? And what assumptions do you need to make?

2.   Consider shaking a thumb tack in a cup and tossing the tack out onto a table. It can land one of two ways (see picture). How would you construct a relative frequency probability for the probability that the tack lands point down?

Can you construct a model-based probability? Justify your answer.

3. A random number table is constructed from a sequence of digits. Each new digit in the sequence is obtained by choosing a digit at random from the 10 digits $0, 1, \ldots, 9$. Say whether each of the following statements is true or false and why.

(a) Each column should have the same number of 9s in it.

(b) Each column should have a similar number of 4s as 5s.

(c) After three 5s in a row, the next number is less likely to be a 5.

(d) We are less likely to see the sequence 1,2,3,4,5 than to see the sequence 2,7,4,9,3.

4.4 Simple Probability Models

4.4.1 Sample spaces

We begin with the idea of a *random experiment*, that is an experiment whose outcome cannot be predicted. The term experiment is used in its widest sense. It can mean either a naturally occurring phenomenon (e.g. measuring the height of high tide on a given day, counting the number of aphids on a leaf), a scientific experiment (e.g. measuring the speed of sound or the blood pressure of a patient), or a sampling experiment (e.g. choosing a person at random from a class of students using the lottery method and recording some characteristic of the person).

> A *sample space*, *S*, for a random experiment is the set of all possible outcomes of the experiment.

In simple examples we can represent the sample space simply as a list. With more complicated examples some mathematical representation may be necessary. There are two important considerations in the way we list the outcomes.

Firstly, every outcome must be represented. Secondly, to avoid ambiguity, no outcome can be represented twice. This means that any outcome gives rise to one and only one member of the list.

You will find that, apart from examples based on data tables and Case Studies, many of the examples in this chapter are very simple and not very "realistic". For example, you will often see examples about tossing a coin, or sampling colored balls from a barrel. However, just as tossing a coin can serve as a useful model for sex outcomes when having children, we shall find in Chapter 5 that these very simple physical experiments will become the basis of models for a vast array of real applications. In the meantime, however, we shall just concentrate on using them to enable us to explore how probabilities and probability models behave.

Example 4.4.1.

(**a**) If we toss a coin twice we can represent the 4 possible outcomes in terms of heads (H) and tails (T) as
$$S = \{HH, HT, TH, TT\},$$
where HT, for example, indicates a head followed by a tail.

(**b**) Similarly 3 tosses give
$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

(**c**) If we roll two dice and record the numbers facing uppermost on each die we could use
$$S = \{(1,1), (1,2), \dots, (1,6), (2,1), \dots, (2,6), (3,1), \dots, (6,6)\},$$
where we have represented each of the 36 possible outcomes as a pair. For example, $(1,6)$ represents rolling a 1 with the first die and a 6 with the second.

(**d**) Suppose we interview a person at random and record their religious preference (if any). A sample space for the outcome of the interview might be
$$S = \{\text{Buddhism, Christianity, Hinduism, Islam, Other, None}\}.$$
Here the category "Other" would be used to capture all of those who adhere to a religion not listed. Those under "None" would consist of people who adhere to no religion. Have we got a sample space? We have catered for all religions. However, we have not listed all answers people will give us. We need a further category which we might call "Nonresponse" to include, for example, those who will refuse to answer. Also, as with most classification systems, there are problems with category definition. Consider Christianity. There may be some sects which you would be unsure about whether to list under Christianity or under Other.

(**e**) Suppose that in contrast to (a) and (b) we now toss a coin until the first tail appears. Then,
$$S = \{T, HT, HHT, HHHT, \dots\}.$$

The rows of dots here means that the pattern keeps on repeating for ever. There is no limit to the number of heads we could conceivably throw before our first tail.

(f) Suppose the experiment is to measure tomorrow's rainfall. A possible S is the set of all numbers greater than or equal to zero which we could write in set notation as

$$S = \{x : x \geq 0\}.$$

The curly brackets designate a set, and the colon stands for "such that". The expression for S is read as "the set of all x such that x is greater than or equal to zero." If we were prepared to believe that there was no way that the day's rainfall would be greater than 30mm, we could restrict S to all numbers from 0mm to 30mm which we could write as

$$S = \{x : 0 \leq x \leq 30\}.$$

∎

These examples illustrate a number of ideas. The first five, (a) to (e), are just ordered lists, with a finite number of elements in (a) to (d), and an infinite number of elements in (e). Such sample spaces are said to be *discrete* in contrast to (f) which is called a *continuous* sample space as the actual rainfall can take any value over an interval.

More that one sample space can be used to describe an experiment. This is why we talk about *a* sample space, and not *the* sample space for an experiment. In (b) above, we could count the number of heads in the three tosses and use $S_1 = \{0 \text{ heads}, 1 \text{ heads}, 2 \text{ heads}, 3 \text{ heads}\}$. Every outcome of the experiment is represented and represented only once, as required. Which sample space we use depends upon the type of question we wish to answer. For example, S_1 lets us talk about the number of heads but does not let us distinguish between the order in which heads and tails fall, while S lets us address both issues.

4.4.2 Events

We often want to talk about a collection of outcomes which share some characteristic, e.g. the outcomes resulting in "at least one head" if we toss a coin twice. This leads to the following definition of an event:

> An *event* is a collection of outcomes.
> An event *occurs* if any outcome making up that event occurs.

The sample space itself is an event. An event may contain only a single outcome.[7]

---

[7]Technically, an event is a subset of the sample space.

Example 4.4.2.

(a) In Example 4.4.1(a), we toss a coin twice, giving $S = \{HH, HT, TH, TT\}$. The event $A =$ "at least one head" is given by $A = \{HH, HT, TH\}$. If any one of these three outcomes occurs when we toss the coin twice, then event A has occurred.

(b) In Example 4.4.1(d), the event $B =$ "has a religious preference which is not Buddhism or Christianity" is given by $B = \{\text{Hinduism, Islam, Other}\}$.

(c) In Example 4.4.1(f), the event $C =$ "rainfall between 5 and 20 mm inclusive" can be written in mathematical notation as the interval of values

$$C = \{x : 5 \leq x \leq 20\}.$$

∎

The ***complement*** of an event A, denoted $\overline{A}$, occurs if A does not occur.

The complement of A, denoted $\overline{A}$, contains all outcomes not in A. It is sometimes helpful to read $\overline{A}$ as "not A".

Example 4.4.3.

(a) The complement of A in Example 4.4.2(a) is $\overline{A} = \{TT\}$. Verbally, the complementary event to "at least one head" is "no heads" which in this case is the same as "two tails".

(b) In Example 4.4.1(b), where we toss a coin three times, the event $B =$ "at least two heads" $= \{HHH, HHT, HTH, THH\}$ has complement $\overline{B} = \{HTT, THT, TTH, TTT\}$. Verbally, the complementary event to "at least two heads" is "at most one head" or, equivalently here, "at least two tails". ∎

It is useful to represent events diagrammatically. These are so-called *Venn diagrams*. We tend to represent the sample space S as a rectangular box. Events inside S are represented by the contents of a closed shape. Any shape will do, although we shall usually use a circle, as in Fig. 4.4.1(a). In Fig. 4.4.1(b) we have shaded the contents of A (which we think of as representing all of the outcomes in A), whereas in Fig. 4.4.1(c) we have shaded the contents of $\overline{A}$.
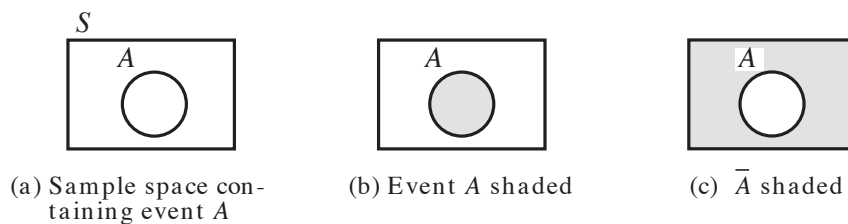
(a) Sample space containing event A (b) Event A shaded   (c) $\overline{A}$ shaded

Figure 4.4.1 :   An event A in the sample space S.

Exercises on Section 4.4.2

1. In Example 4.4.1(c), let $A =$ "sum of the faces uppermost is 4". List the outcomes in A.

2. In Example 4.4.1(e), Let $A =$ "even number of tosses before the tail". What outcomes are in A? Describe $\overline{A}$ both verbally and by listing its outcomes.

3. In Example 4.4.1(f), let $A =$ "rainfall no more that 2mm". Describe A mathematically. Describe $\overline{A}$, the complement of A both verbally and mathematically.

4. Write down a sample space for tossing a coin until we have two tails, three heads, or a maximum of four tosses. What outcomes are in the event $A =$ "3 tosses made"?

4.4.3 Combining events

The set theory notations of union ($\cup$) and intersection ($\cap$) provide a useful shorthand for writing expressions involving events. For two events A and B, $A \cup B$ represents "A or B occurs" (where "or" is used in the inclusive sense of A or B or both[8]), whereas $A \cap B$ represents "both A and B occur". For part of this chapter we shall use both words and symbols to remind the reader of their relationships.

$\boldsymbol{A} \cup \boldsymbol{B}$ contains all outcomes in A **or** B (or both).

$\boldsymbol{A} \cap \boldsymbol{B}$ contains all outcomes which are in **both** A and B.

In practice, therefore, we read "$\cup$" as "or" (in the inclusive sense) and "$\cap$" as "and".

(a) Events A and B (b) "A or B" shaded   (c) "A and B" shaded   (d) Mutually exclusive events

Figure 4.4.2 :   Two events.

Example 4.4.4.

(a) In Example 4.4.1(a) we have $S = \{HH, HT, TH, TT\}$. Let A be the event "at least one head" and B be the event "at least one tail". Then

$$A = \{HH, HT, TH\}, \quad B = \{HT, TH, TT\},$$
$$A \text{ and } B = A \cap B = \{HT, TH\} = \text{"exactly one head"}, \quad \text{and}$$
$$A \text{ or } B = A \cup B = \{HH, HT, TH, TT\} = S.$$

---

[8]This is sometimes written as "and/or". Another phrase we shall also use here is "at least one of A and B occurs".

(b) In the rainfall example [Example 4.4.1(f)] let A be the event "at least 10mm of rain" and let B be the event "between 5mm and 15mm inclusive". Then

$$A = \{x : x \geq 10\}, \quad B = \{x : 5 \leq x \leq 15\},$$
$$A \text{ and } B = A \cap B = \{x : 10 \leq x \leq 15\}, \text{ and}$$
$$A \text{ or } B = A \cup B = \{x : x \geq 5\}.$$

∎

Two events A and B which have no outcomes in common are said to be ***mutually exclusive***.

Mutually exclusive events cannot occur at the same time.

You may find the phrase, "A excludes B" a useful memory aid for the meaning of "mutually exclusive". Any event A and its complement $\overline{A}$ are mutually exclusive. We usually represent mutually exclusive events diagrammatically by nonoverlapping shapes as in Fig. 4.4.2(d).

It is conventional to use $\boldsymbol{A \cap B = \varnothing}$ as a shorthand notation for "A and B are mutually exclusive". Here we have introduced an artificial event denoted by $\varnothing$. This is called the ***empty*** or ***null event*** [9] and contains no outcomes.
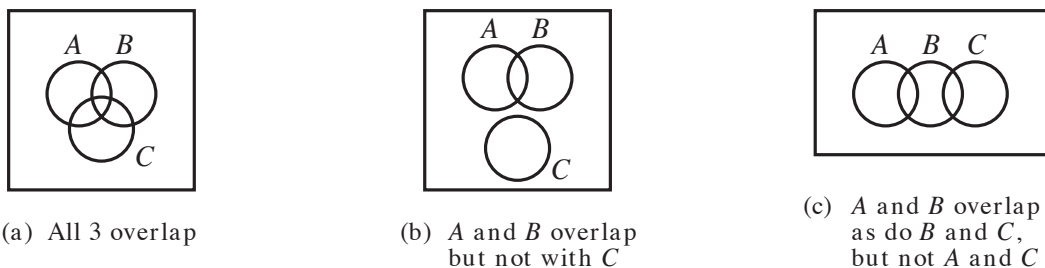
Example 4.4.5.

(a) When tossing a coin twice [Examples 4.4.1(a), 4.4.2(a)] the events "head first toss" $= \{HH, HT\}$ and "tail first toss" $= \{TH, TT\}$ are mutually exclusive. However the two events "head first toss" $= \{HH, HT\}$ and "head second toss" $= \{HH, TH\}$ are not mutually exclusive. Both of these events occur if we observe 2 heads (i.e. $\{HH, HT\} \cap \{HH, TH\} = \{HH\}$).

(b) Consider rolling a die twice [Example 4.4.1(c)]. Let $A = $ "sum from the two faces is 4", $B = $ "3 on first roll" and $C = $ "sum from the two faces is 7". Then $A = \{(1, 3), (2, 2), (3, 1)\}$, $B = \{(3, 1), (3, 2), \ldots, (3, 6)\}$ and $C = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$. Now A and C are mutually exclusive, i.e. $A \cap C = \varnothing$. However A and B can occur together as $A \cap B = \{(3, 1)\}$. Also B and C can occur together as $B \cap C = \{(3, 4)\}$. Fig. 4.4.3(c) shows a situation like this.

∎

We can use Venn diagrams for three or more events as in Fig. 4.4.3.

---

[9]The null event corresponds to the empty set in set theory.

(a) All 3 overlap

(b) *A* and *B* overlap but not with *C*

(c) *A* and *B* overlap as do *B* and *C*, but not *A* and *C*

Figure 4.4.3 : Three events.

In Fig. 4.4.3(a) all of the three events overlap. In Fig. 4.4.3(b) events *A* and *B* have outcomes in common, but event *C* shares no outcomes with either *A* or *B*.

Exercises on Section 4.4.3

1. An experiment consists of tossing a coin and rolling a die. Give a sample space for the experiment. Let $A = $ "die scores 3" and $B = $ "coin is heads". By listing outcomes, write down expressions for $A, B, A \cap B$ (*A* and *B*) and $A \cup B$ (*A* or *B*).

2. For the *ABO* blood system a person can be one of the four phenotypes *A*, *B*, *O* and *AB*. Two people are chosen at random. Give a sample space for the pair of phenotype outcomes. By listing outcomes, give expressions for the following events:
  (a) $C = $ "both people have the same phenotypes";
  (b) $D = $ "at least one person has phenotype *A*"; and
  (c) $C \cap D$ (*C* and *D*).

3. Represent each of the following on a (separate) Venn diagram and then express them in terms of *A* and *B*, using intersections, unions and complements as required:
  (a) *A* occurs but *B* does not;   **(b)** *B* occurs but *A* does not;
  (c) *B* occurs or *A* does not;   **(d)** at least one of *A* or *B* occurs;
  (e) exactly one of *A* and *B* occurs;   **(f)** neither *A* nor *B* occurs (try and write this one down in two different ways).
  *(g) If $A \cap B = \varnothing$, what can we say about $A \cap \overline{B}$ and $\overline{A} \cap B$?

4. Suppose that we choose one month from the 12 months of the year at random.
  (a) Write down a sample space for this experiment.
  Consider the events $A = $ "first 2 months of the year", $B = $ "a month beginning with the letter J" and $C = $ "the last 6 months of the year".
  (b) Construct a Venn Diagram with these three events on it (cf. Fig. 4.4.3) and place the outcomes of the experiment (i.e. the months of the year) in the relevant parts of the diagram.
  (c) Which pairs of events are mutually exclusive and which are not?

(**d**) What outcomes are in: (**i**) $\overline{B}$ (**ii**) B or C ($B \cup C$) (**iii**) B and C ($B \cap C$) (**iv**) $\overline{A}$ or B ($\overline{A} \cup B$) (**v**) A and $\overline{B}$ ($A \cap \overline{B}$)?

4.4.4 Probability distributions

Traditional usage dictates that probabilities are numbers scaled to lie between zero and one (or 0% and 100%) and that outcomes with probability zero cannot occur. In addition, we say that events with probability one or 100% are certain to occur. We now go on to define the term *probability distribution*, but only for models with finite sample spaces or with infinite sample spaces that can be represented as a list,[10] e.g. $S = \{H, TH, TTH, TTTH, \dots\}$.

Suppose $S = \{s_1, s_2, s_3, \dots\}$ is such a sample space. A list of numbers $p_1, p_2, \dots$ is a **probability distribution** for S provided the p_i's satisfy both

(**i**) the p_i's lie between zero and one, $\qquad$ ($\mathbf{0 \leq p_i \leq 1}$),

and (**ii**) the sum of all the p_i's is one. $\qquad$ ($\mathbf{p_1 + p_2 + \dots = 1}$).

According to the probability model, p_i is the probability that outcome s_i occurs. We write $p_i = \mathrm{pr}(s_i)$.

> Probabilities lie between 0 and 1, and they add to 1.

In practice our aim is not just to specify a mathematically valid probability model, namely one that satisfies conditions (i) and (ii) above, but also to specify a model in which the stated probabilities give a good approximation to the actual behavior of the experiment.

Example 4.4.6.

(**a**) Consider tossing a coin twice so that $S = \{HH, HT, TH, TT\}$. For a fair coin, each outcome should be equally likely so that the probabilities for the four outcomes, p_1, p_2, p_3 and p_4, should be identical. If we want them to add to 1 each value must therefore be $\frac{1}{4}$, i.e.

$$\mathrm{pr}(HH) = \tfrac{1}{4}, \quad \mathrm{pr}(HT) = \tfrac{1}{4}, \quad \mathrm{pr}(TH) = \tfrac{1}{4}, \quad \mathrm{pr}(TT) = \tfrac{1}{4}.$$

These probabilities constitute a probability distribution for S.

(**b**) Consider choosing a three child family at random and looking at the sexes of the children from first born to last born, then

$$S = \{GGG, GGB, GBG, BGG, GBB, BGB, BBG, BBB\}.$$

Let us assume that looking at the sexes of a randomly chosen three-child family is like tossing a fair coin three times and that each of the 8 (=

---

[10]Technically such a sample space is said to be countably infinite.

2^3) outcomes in S is equally likely to occur. Since the probabilities must add to 1, each outcome has probability $\frac{1}{8}$.

(c) To vary this a little, consider a couple producing children. They will stop when they have a child of each sex, or stop when they have 3 children. Now for this "experiment" the sample space is

$$S = \{GGG, GGB, GB, BG, BBG, BBB\}.$$

By comparing (a) and (b), a reasonable probability distribution might be

$$\text{pr}(GGG) = \text{pr}(GGB) = \text{pr}(BBG) = \text{pr}(BBB) = \tfrac{1}{8},$$
and $\qquad \text{pr}(GB) = \text{pr}(BG) = \tfrac{1}{4}.$

These values are between 0 and 1 and add to 1, so that they qualify as a probability distribution.[11]

(d) Frequently a sample space consists of just a list of numbers. For example, if the outcome of the "experiment" in (b) is the number of girls, we have $S = \{0, 1, 2, 3\}$. Then

$$\text{pr}(0) = \text{pr}(BBB) = \tfrac{1}{8},$$
$$\text{pr}(1) = \text{pr}(GBB) + \text{pr}(BGB) + \text{pr}(BBG) = \tfrac{3}{8},$$
$$\text{pr}(2) = \text{pr}(GGB) + \text{pr}(GBG) + \text{pr}(BGG) = \tfrac{3}{8},$$
and $\qquad \text{pr}(3) = \text{pr}(GGG) = \tfrac{1}{8}.$

This gives us the probability distribution associated with S. We can check our arithmetic by noting that the above four probabilities add up to 1. This example looks ahead to Chapter 5 where we discuss discrete random variables. Here the number of girls is called a (discrete) random variable.

Probabilities of events

> The ***probability of event A*** can be obtained by adding up the probabilities of all the outcomes in A.

---

[11] After reading Section 4.7, you will be able to derive these probabilities as following from the assumptions that $\text{pr}(G) = \text{pr}(B) = \tfrac{1}{2}$ and the sexes of different children are statistically independent.

Example 4.4.7.

(**a**) In Example 4.4.6(a), let the event $A = $ "at least one head" $= \{HH, HT, TH\}$. By adding the probabilities of the 3 outcomes in A we find

$$\mathrm{pr}(A) \; = \; \mathrm{pr}(HH) \; + \; \mathrm{pr}(HT) \; + \; \mathrm{pr}(TH) = \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{3}{4}.$$

(**b**) Similarly, for Example 4.4.6(c) with C being the event "first child is a girl", we have $C = \{GGG, GGB, GB\}$ and

$$\mathrm{pr}(C) = \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{4} = \tfrac{1}{2}.$$

Equally likely outcomes

If S consists of 10 equally likely outcomes, each with probability p then, since the probabilities add to one, we have $10p = 1$ or $p = \tfrac{1}{10}$. Thus each outcome has a probability of $\tfrac{1}{10}$. If A has four outcomes in it, then using our addition rule we have $\mathrm{pr}(A) = \tfrac{1}{10} + \tfrac{1}{10} + \tfrac{1}{10} + \tfrac{1}{10} = \tfrac{4}{10}$. More generally, for any finite sample space with equally likely outcomes,

$$\mathbf{pr(A)} \; = \; \frac{\textbf{Number of outcomes in } \; A}{\textbf{Total number of outcomes in } S}.$$

Example 4.4.8 In Example 4.4.7(a), 3 out of 4 outcomes are in A, so that $\mathrm{pr}(A) = \tfrac{3}{4}$.

Example 4.4.9 Table 4.4.1, called a ***two-way frequency table*** or *contingency table*, cross classifies job losses in the US over a three year period. Job losses are broken down by the gender of the person who lost the job and the reason given for losing it. The entries in the table represent the number of job losses (in thousands) by people of the particular gender for a particular reason. There were 5,584,000 jobs lost (to the nearest thousand) and of these, 1,703,000 were lost by males because the workplace moved or closed down.

Table 4.4.1 : Job Losses in the US (in thousands) for 1987 to 1991[a]

Lay off Numbers (thousands)	Reason for Job Loss			Total
	Workplace moved/closed	Slack work	Position abolished	
Male	1,703	1,196	548	3,447
Female	1,210	564	363	2,137
Total	2,913	1,760	911	5,584

[a]Source: Constructed from data in The World Almanac, [1993, p. 157].

Suppose we decided to choose a lost job at random so that we could investigate the circumstances. The sample space for this experiment consists of all

the $5,584,000$ lost jobs. Since these are all equally likely to be chosen, we can obtain probabilities by counting numbers of outcomes. The outcomes making up the event "lost by female as workplace moved/closed down" are all job losses with this property and there are $1,210,000$ of them. Thus,

$$\text{pr}(\textit{lost by female as workplace moved/closed down}) = \frac{1,210,000}{5,584,000} = 0.2167$$

to 4 decimal places. There were $2,137,000$ jobs lost by females, so that

$$\text{pr}(\textit{lost by female}) = \frac{2,137,000}{5,584,000} = 0.3827\,.$$

The event "lost by male for slack work" has 1,196,000 outcomes, so

$$\text{pr}(\textit{lost by male for slack work}) = \frac{1,196,000}{5,584,000} = 0.2142\,.$$

We can convert these to percentages by multiplying by 100% so that the last answer becomes 21.42%.

The set of 6 combinations of classes in the table ("lost by male because workspace moved/closed", ..., "lost by female as position abolished") forms an alternative sample space for this experiment as all eventualities are allowed for and no "outcomes" are represented twice. However, the sample space consisting of all 5,584,000 individual job losses was more useful because, since its outcomes are all equally likely, we could obtain any relevant probabilities almost immediately. Note that the probabilities relating to entries in the table turned out to be the corresponding proportions of the population of job losses. Having established that connection, the set of 6 classes represented in Table 4.4.1 also functions as a usable sample space for the experiment because we can write down the probabilities for its 6 outcomes. The 2 genders, or the 3 reasons for job loss, are yet more possible sample spaces – although these latter sample spaces do not allow us to consider questions about the relationship between gender and reason for job loss.

Exercises on Section 4.4.4

1. Two dice are rolled. What is the probability that the sum of the two faces uppermost is (i) 9   (ii) even?

2. (Example 4.4.9 revisited).
  (a) What is the probability that a randomly chosen job loss was: **(i)** by a female whose position was abolished?   **(ii)** caused by the position being abolished?

  (b) Take the 6 classes in the table and represent them as a sample space with an associated probability distribution.

(c) Do the same thing as (b) but using the three categories of job loss as a sample space.

(d) In Example 4.4.9, why did we employ all job losses as our sample space and not the sample space in (b)?

3. (Background thinking about the data in Table 4.4.1)

(a) In Example 4.4.9, we were careful not to equate the 5,584,000 job losses over the 3 year period with 5,584,000 people. Why are the two not equivalent?

(b) There were fewer job losses by females over this period. Does this demonstrate that women are more reliable workers?

(c) If you wanted to compare the relative chance of a random male losing his job to that of a random female losing her job, what measure would you use?

4.4.5 Probabilities and proportions

When we have a real (finite) population of units (e.g. people), the theory of equally likely outcomes tells us that, when we choose a unit at random, the probability that a unit with property A is chosen is numerically identical to the proportion of units in the population with property A. For example, if 10% of the population is left-handed, the chances that a randomly chosen person is left-handed is also 10%.

The concepts of a proportion and a probability are quite distinct. A proportion is a partial description of a real population – a form of summary. Probabilities tell us about the chances of something happening in a random experiment. However, the fact that proportions are numerically identical to probabilities for a real population under the experiment "choose a unit at random" means that we can use the probability notation and any formulae derived for manipulating probabilities to solve problems involving proportions as well. At times, where it makes good practical sense, we will express a real problem in terms of proportions of a population rather than probabilities. We prefer to do this then rather than introducing the artificial "choose a unit at random" that would enable us to write everything in terms of probabilities.

Quiz on Section 4.4

1. What is a sample space? What are the two essential criteria that must be satisfied by a possible sample space? (Section 4.4.1)

2. What is an event? (Section 4.4.2)

3. If A is an event, what do we mean by its complement $\overline{A}$? When does $\overline{A}$ occur? (Section 4.4.2)

4. If A and B are events, when does A or B $(A \cup B)$ occur? When does A and B $(A \cap B)$ occur? Using Venn diagrams, show what outcomes are in A or B and what outcomes are in A and B? Do the outcomes in A or B include any outcomes in A and B? (Section 4.4.3)

5. How do we denote the fact that events C and D cannot both occur? What adjective is used to describe such events? (Section 4.4.3)

6. What are the two essential properties of a probability distribution $p_1, p_2, \ldots p_n$? (Section 4.4.4)
7. How do we get the probability of an event from the probabilities of outcomes that make up that event? (Section 4.4.4)
8. If all outcomes are equally likely, how do we calculate $\mathrm{pr}(A)$? (Section 4.4.4)
9. How do the concepts of a proportion and a probability differ? Under what circumstances are they numerically identical? What does this imply about using probability formulae to manipulate proportions of a population? (Section 4.4.5)

4.5 Probability Rules

4.5.1 One or two events

The following rules are clear for finite sample spaces[12] in which we obtain the probability of a event by adding the p_i's for all outcomes in that event. The outcomes do not need to be equally likely.

Rule 1: The sample space is certain to occur.[13]

$$\mathbf{pr}(S) = 1 \ .$$

Rule 2: $\mathrm{pr}(A \text{ does } \boldsymbol{not} \text{ occur}) = 1 - \mathrm{pr}(A \text{ does occur})$.

$$\mathbf{pr}(\overline{A}) = 1 - \mathbf{pr}(A) \ .$$

[Alternatively, pr(An event occurs) = 1 - pr(it doesn't).]

*Rule 3[14]**:** $\mathrm{pr}(A \text{ or } B \text{ occurs}) = \mathrm{pr}(A \text{ occurs}) + \mathrm{pr}(B \text{ occurs}) - \mathrm{pr}(\text{both occur})$.

$$\mathbf{pr}(A \cup B) = \mathbf{pr}(A) + \mathbf{pr}(B) - \mathbf{pr}(A \cap B) \ .$$

[In adding $\mathrm{pr}(A) + \mathrm{pr}(B)$ we use the p_i's relating to outcomes in $A \cap B$ twice. Thus we adjust by subtracting $\mathrm{pr}(A \cap B)$.]

Rule 4: $\mathrm{pr}(A \text{ occurs}) = \mathrm{pr}(A \text{ occurs with } B) + \mathrm{pr}(A \text{ occurs without } B)$.
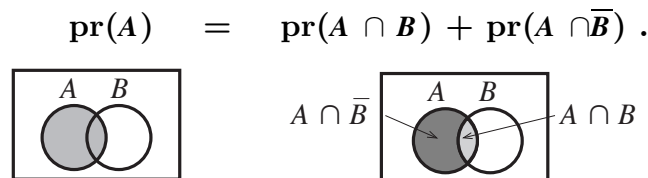
$$\mathbf{pr}(A) \quad = \quad \mathbf{pr}(A \cap B) + \mathbf{pr}(A \cap \overline{B}) \ .$$
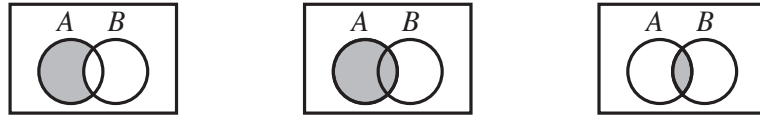

Rearranging Rule 4 we get

---

[12]However the rules apply generally, and not just to finite sample spaces.
[13]The sample space contains all possible outcomes.
[14]This rule has little use later on and may be omitted at first reading.

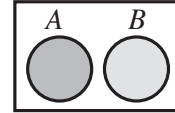$$\mathrm{pr}(A \cap \overline{B}) \quad = \quad \mathrm{pr}(A) \quad - \quad \mathrm{pr}(A \cap B).$$

We can write down these rules from the diagrams without having to remember them.

Mutually exclusive events

If A and B are mutually exclusive, they have no outcomes in common and cannot occur at the same time. Thus $\mathrm{pr}(A \cap B) = 0$ and

$$\mathbf{pr}(A \cup B) = \mathbf{pr}(A) + \mathbf{pr}(B).$$

This is a special case of Rule 3.

Example 4.5.1. A random number from 1 to 10 is selected from a table of random numbers. Let A be the event that "the number selected is 9 or less." As all 10 possible outcomes are equally likely, $\mathrm{pr}(A) = \frac{9}{10} = 0.9$. The complement of event A has only one outcome, so $\mathrm{pr}(\overline{A}) = \frac{1}{10} = 0.1$. Note that

$$\mathrm{pr}(A) = 1 - \mathrm{pr}(\overline{A}) = 1 - 0.1 = 0.9 \,.$$

This formula is very useful in any situation where $\mathrm{pr}(\overline{A})$ is easier to obtain than $\mathrm{pr}(A)$. It is used in this way in Example 4.5.4 and frequently thereafter.

*Example 4.5.2. Suppose that between the hours of 9.00am and 5.30pm, Dr Wild is available for student help 70% of the time, Dr Seber is available 60% of the time, and both are available (simultaneously) 50% of the time. A student comes for help at some random time in the above hours. Let $A =$ "Wild in" and $B =$ "Seber in". Then $A \cap B =$ "both in". If a time is chosen at random, then $\mathrm{pr}(A) = 0.7$, $\mathrm{pr}(B) = 0.6$ and $\mathrm{pr}(A \cap B) = 0.5$.

What is the probability that at least one of the two Professors is available? The event of interest is $A \cup B$ and, by Rule 3,

$$\begin{aligned}
\mathrm{pr}(A \cup B) &= \mathrm{pr}(A) \ + \ \mathrm{pr}(B) \ - \ \mathrm{pr}(A \cap B) \\
&= 0.7 \ + \ 0.6 \ - \ 0.5 \ = \ 0.8.
\end{aligned}$$

What is the probability that only Wild is available? The event of interest is $A \cap \overline{B}$. Using the rearranged version of Rule 4,

$$\mathrm{pr}(A \cap \overline{B}) = \mathrm{pr}(A) - \mathrm{pr}(A \cap B) = 0.7 - 0.5 = 0.2 \,.$$

Finally, what is the probability that neither is available? The event of interest is $\overline{A} \cap \overline{B}$ but it is easy to see (e.g. from a diagram or from the verbal description) that this event is the complement of $A \cup B$ so that by Rule 2

$$\mathrm{pr}(\overline{A} \cap \overline{B}) = 1 - \mathrm{pr}(A \cup B) = 0.2\,.$$

Exercises on Section 4.5.1

1. A young man with a barren love life feels tempted to become a contestant on the television game show "Blind Date". He decides to watch a few programs first to assess his chances of being paired with a suitable date, namely someone he finds attractive and no taller than he is (as he is hung up about his height). After watching 50 female contestants, he decides that he is not attracted to 8, that 12 are too tall, and 16 are either unattractive or too tall (or both). If these figures are typical, what is the probability of getting someone:

(**a**) who is both unattractive and too tall?

(**b**) whom he likes i.e. is not unattractive or too tall?

(**c**) who is too tall but not unattractive?

2. A house needs to be reroofed during Spring. To do this a dry, windless day is needed. The probability of getting a dry day is 0.7, a windy day is 0.4 and a wet, windy day is 0.2. What is the probability of getting:

(**a**) a wet day?

(**b**) a day which is either wet or windy, or both?

(**c**) a day when the house can be reroofed?

3. For the data and situation in Example 4.4.9, what is the probability that a random job loss:

(**a**) was not by a male who lost it because the workplace moved?

(**b**) was by a male or someone who lost it because the workplace moved?

(**c**) was by a male but for some reason other than the workplace moving?

*4. Suppose A and B are mutually exclusive events with $\mathrm{pr}(A) = 0.3$ and $\mathrm{pr}(B) = 0.4$. Find
(**a**) $\mathrm{pr}(\overline{A})$ (**b**) $\mathrm{pr}(A \cap B)$ (**c**) $\mathrm{pr}(A \cup B)$ (**d**) $\mathrm{pr}(\overline{A} \cap B)$.

*5. Try to convince yourself of the correctness of Rule 3 by drawing diagrams.

4.5.2 More than two events

We can write "at least one of the k events A_1, A_2, ... , A_k occurs" as "A_1 or A_2 or ... or A_k occurs (in symbols $A_1 \cup A_2 \cup \cdots \cup A_k$)". Similarly, we can write "every one of the k events A_1, ... , A_k occurs" as "A_1 and A_2 and ... and A_k occurs (in symbols $A_1 \cap A_2 \cap \cdots \cap A_k$)".

Events $A_1, A_2, \ldots, A_k$ are all mutually exclusive if they have no overlap, i.e. if no two of them can occur at the same time. If $A_1, \ldots, A_k$ are mutually exclusive then we can get the probability of their union simply by adding, i.e.,

$$\mathrm{pr}(A_1 \cup A_2 \cup \ldots \cup A_k) \; = \; \mathrm{pr}(A_1) \; + \; \mathrm{pr}(A_2) + \ldots \; + \; \mathrm{pr}(A_k)$$

$$\left(\; = \sum_{i=1}^{k} \mathrm{pr}(A_i) \right).$$

Partitions: Events $C_1, C_2, \ldots, C_k$ form a partition of the sample space if they are mutually exclusive and together account for all possible outcomes (i.e. $C_1 \cup C_2 \cup \ldots \cup C_k = S$). Any event and its complement (e.g. A and $\overline{A}$) form a two-event partition. We have a pictorial representation of a partition of 5 events in Fig. 4.5.1. As a concrete example, a jigsaw puzzle represents a partition of a picture.
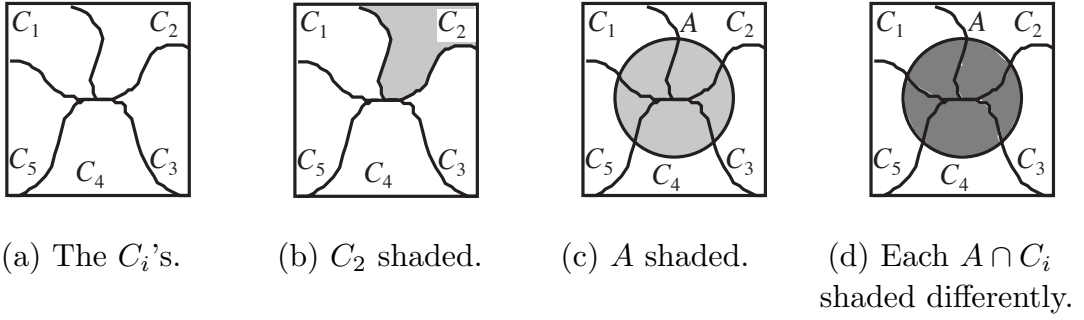

(a) The C_i's.   (b) C_2 shaded.   (c) A shaded.   (d) Each $A \cap C_i$ shaded differently.

Figure 4.5.1 :   Partition theorem.

> A partition is a way of dividing up
> a sample space into separate pieces.

If we have such a partition, then adding up the probabilities associated with the (mutually exclusive) shaded bits in Fig. 4.5.1(d) gives us Rule 5.

Rule 5: (Partition Theorem)

$$\mathrm{pr}(A) = \mathrm{pr}(A \text{ and } C_1) + \mathrm{pr}(A \text{ and } C_2) + \ldots + \mathrm{pr}(A \text{ and } C_k)$$
$$= \mathrm{pr}(A \cap C_1) + \mathrm{pr}(A \cap C_2) + \ldots \mathrm{pr}(A \cap C_k)$$
$$\left(\; = \sum_{i=1}^{k} \mathrm{pr}(A \cap C_i) \right).$$

This is a generalization of Rule 4 where our partition was $C_1 = B$ and $C_2 = \overline{B}$.

Example 4.5.3 A number is drawn at random from 1 to 10 so that the sample space $S = \{1, 2, ..., 10\}$. Let A = "an even number chosen" and let $C_1 = \{1, 2, 3\}$, $C_2 = \{4, 5, 6\}$ and $C_3 = \{7, 8, 9, 10\}$ be our partition of S. We shall now verify rule 5 in this case for this situation.

Since outcomes are equally likely, $\text{pr}(A) = \frac{5}{10} = 0.5$. Also $A \cap C_1 = \{2\}$, $A \cap C_2 = \{4, 6\}$ and $A \cap C_3 = \{8, 10\}$ so that

$$\text{pr}(A \cap C_1) + \text{pr}(A \cap C_2) + \text{pr}(A \cap C_3) = \frac{1}{10} + \frac{2}{10} + \frac{2}{10} = \frac{5}{10} = \text{pr}(A).$$

Example 4.5.4 Table 4.5.1, called a ***two-way table of proportions*** or ***contingency table*** (introduced briefly in Section 3.3), cross classifies couples in the US, who are not married to each other, by the marital status of the male and female partners. Each entry within the table is the proportion of couples with a given combination of marital statuses. Let us consider choosing a couple at random. We shall take all of the couples represented in the Table as our sample space. The table proportions give us the probabilities of the events defined by the row and column titles (see Example 4.4.9 and Section 4.4.5). Thus, the probability of getting a couple where the male has never been married and the female member is divorced is 0.111. Here, "married to other" means living as a member of a couple with one person while married to someone else.

Table 4.5.1 : Proportions of Unmarried Male-Female Couples Sharing a Household in the US, 1991[a]

Male	Female				Total
	Never Married	Divorced	Widowed	Married to other	
Never Married	0.401	.111	.017	.025	.554
Divorced	.117	.195	.024	.017	.353
Widowed	.006	.008	.016	.001	.031
Married to other	.021	.022	.003	.016	.062
Total	.545	.336	.060	.059	1.000

[a]Source: Constructed from data in The World Almanac, [1993, p. 942].

The 16 cells in the table, which relate to different combinations of the status of the male and the female partner, correspond to mutually exclusive events (no couple belongs to more than one cell). For this reason, we can obtain probabilities of unions by adding cell probabilities. Moreover, these 16 mutually exclusive events account for all couples so they form a partition of the sample space. The 4 events classifying the status of the female alone form another partition. So too do the 4 events relating to classifying the male alone.

Each row total in the table tells us about the proportion of couples with males in the given category regardless of the status of the females. Thus, in

35.3% of couples, the male is divorced, or equivalently, 0.353 is the probability that the male of a randomly chosen couple is divorced. Use of a row total in the table in this way is an illustration of the Partition Theorem in action. The event A of interest here is "male divorced". The 4 partitioning events are $C_1 =$ "female never married", ... , $C_4 =$, "female married to other". Adding along the row corresponds to $\sum \text{pr}(A \text{ and } C_i)$ which by the Partition Theorem gives us $\text{pr}(A)$, or the probability that the male is divorced. The column totals work the same way for females.

We shall continue to use Table 4.5.1 to further illustrate the use of the probability rules in the previous subsection. The complement of the event "at least one member of the couple has been married" is the event that both are in the "never been married" category. Thus by Rule 2,

$$\text{pr}(\textit{At least one has been married}) = 1 - \text{pr}(\textit{Both never married})$$
$$= 1 - 0.401 = 0.599 \,.$$

Note how much simpler it was in this case to use the complement than to calculate the desired probability directly by adding the probabilities of the 15 cells which fall into the event "at least one married".

Exercises on Section 4.5.2

1. Do the events A, B, and C listed in problem 4 of Exercises 4.4.3 form a partition of the sample space? Why or why not?

2. Chlorofluorocarbons (CFCs) have been identified as important causes of the depletion of the ozone layer. Of the 750,000 metric tons of these substances used worldwide in 1991 (*TIME*, 17 February 1992), 15% were used in aerosol sprays, 15% in refrigeration, 20% in vehicle air-conditioning, 24% in cleaning fluids and 24% in foam (for insulation, packing etc.). Does this form a partition of CFC usage? Justify your answer.

3. Using the data in Table 4.5.1, what is the probability that for a randomly chosen unmarried couple:

  (a) the male is divorced or married to someone else?

  (b) both the male and the female are either divorced or married to someone else?

  (c) neither is married to anyone else?

  (d) at least one is married to someone else?

  (e) the male is married to someone else or the female is divorced or both?

  (f) the female is divorced and the male is not divorced?

  (g) Show how the column sum which gives pr(female is divorced) $= 0.336$ is an example of the Partition Theorem.

*1. Why in the formula $\text{pr}(A \cup B) = \text{pr}(A) + \text{pr}(B) - \text{pr}(A \cap B)$ do we subtract $\text{pr}(A \cap B)$? (Section 4.5.1)

2. If A and B are mutually exclusive, what is the probability that both occur? What is the probability that at least one occurs (i.e. that $A \cup B$) occurs? (Section 4.5.1)

3. How do we find the probability of a union of two or more mutually exclusive events? (Section 4.5.2)

4. What does it mean for events $C_1, C_2, \ldots, C_k$ to be a partition of the sample space? If $k = 2$, how is C_2 related to C_1? (Section 4.5.2)

4.6 Conditional Probability

4.6.1 Definition

Our assessment of the chances that an event will occur can be very different depending upon the information that we have. An estimate of the probability that your house will collapse tomorrow should clearly be much larger if a violent earthquake was expected than it would be if there were no reason to expect unusual seismic activity.

The two examples which follow give a more concrete demonstration of how an assessment of the chances of an event A occurring may change radically if we are given information about whether event B has occurred or not.

Example 4.6.1. Suppose we toss two fair coins and $S = \{HH, HT, TH, TT\}$. Let $A =$ "two tails" $= \{TT\}$ and $B =$ "at least one head" $= \{HH, HT, TH\}$. Since all four outcomes in S are equally likely, $P(A) = \frac{1}{4}$. However, if we know that B has occurred, then A cannot occur. Hence the *conditional* probability of A *given that B has occurred* is zero.

Example 4.6.2. Table 4.6.1 was obtained by cross-classifying 400 patients with a form of skin cancer, called malignant melanoma, with respect to the histological type[15] of their cancer and its location (site) on their bodies. We see, for example, that 33 patients have nodular melanoma on the trunk while a total of 226 have some form of melanoma on the extremities. Suppose we were to select one of the 400 patients at random.

---

[15]Histological type means the type of abnormality observed in the cells that make up the cancer.

Table 4.6.1 : Four Hundred Melanoma Patients by Type and Site[a]

Type	Site			Row Totals
	Head and Neck	Trunk	Extremities	
Hutchinson's melanomic freckle	22	2	10	34
Superficial spreading melanoma	16	54	115	185
Nodular	19	33	73	125
Indeterminant	11	17	28	56
Column Totals	68	106	226	400

[a]Reproduced from Plackett [1981].

Let A = "cancer is on the Trunk". Clearly, $\text{pr}(A) = 106/400$. Let B be the event "cancer type is Nodular". If we are now told that a patient with Nodular cancer was selected, then the probability that the selected patient has cancer on the trunk may be different – this latter probability is $\text{pr}(A \text{ given } B)$. We are now only concerned with the patients who have nodular cancer, i.e. the 125 patients in the third row of the table. Of this group, 33 have cancer on the trunk and each of them is equally likely to be chosen from the group with nodular cancer. Hence, (where "#" is read "number of")

$$\text{pr}(A \text{ given } B) = \frac{33}{125} = \frac{\# \text{ Nodular patients with cancer on Trunk}}{\# \text{ Nodular patients}},$$

and we denote this probability by $\text{pr}(A \mid B)$. In more general terms

$$\text{pr}(A \mid B) = \frac{\# \text{ outcomes in } A \text{ and } B}{\# \text{ outcomes in } B}$$
$$= \frac{\# \text{ outcomes in } A \text{ and } B / \# \text{ outcomes in } S}{\# \text{ outcomes in } B / \# \text{ outcomes in } S} = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}.$$

What about sample spaces where the outcomes are not all equally likely? The above expression for $\text{pr}(A \mid B)$ can still be justified in much the same way when probabilities are regarded as long run relative frequencies. We therefore use the expression as a general definition of $\text{pr}(A \mid B)$ for all situations.

> **The (conditional) probability of *A* occurring given that *B* occurs is given by**
> $$\text{pr}(A \mid B) = \frac{\text{pr}(A \cap B)}{\text{pr}(B)} \ .$$

Hence the probability that A occurs given B has occurred is the probability that both occur divided by the probability B occurs.[16]

Similarly,
$$\text{pr}(B \mid A) = \frac{\text{pr}(A \cap B)}{\text{pr}(A)}.$$

Example 4.6.3. The data in Table 4.6.2 comes a telephone poll of 800 adult Americans carried out in 1993. The question asked was: "Should smoking be banned from workplaces, should there be special smoking areas, or should there be no restrictions?"

Table 4.6.2 : Proportions of Smokers and Non-smokers and Their Responses to Restrictions[a]

	Banned	Special areas	No restrictions	Total
Non-smokers	.3350	.3975	.0238	.7563
Smokers	.0200	.1963	.0274	.2437
Total	.3550	.5938	.0512	1.0000

[a]Source: The results of a telephone poll of 800 adult Americans from *TIME* 18 April 1994.

Suppose that one of the 800 who responded was chosen at random and we want to calculate the (conditional) probability that a person favors banning smoking, given we know whether they smoke or not. We can work these out using the conditional probability formula as follows:

$$\text{pr}(banned \mid non\text{-}smoker) = \frac{\text{pr}(banned \text{ and } non\text{-}smoker)}{\text{pr}(non\text{-}smoker)}$$
$$= \frac{0.3350}{0.7563} = 0.4429.$$

Similarly
$$\text{pr}(banned \mid smoker) = \frac{0.0200}{0.2437} = 0.0821.$$

Note how different these two probabilities are. The probability that a person chooses the category "banned" depends very strongly on whether the person smokes or not, as might be expected. Notice that both of the probabilities are calculated by dividing each entry in column 1 by the corresponding row total. For example, the first probability is simply the proportion of the first row total in the category "banned". It may help the reader to imagine all the numbers multiplied by 10,000, thus removing the decimal points. We can then think in terms of ratios of frequencies rather than ratios of proportions; that is, of 7563 non-smokers, 3350 wanted smoking banned. Our answers can

---

[16]We must have $\text{pr}(B)$ positive. It makes no intuitive sense, anyway, to compute the conditional probability of A given B when B cannot occur.

also be interpreted as proportions rather than probabilities: the proportion of non-smokers in the survey who prefer banning smoking in workplaces is 0.4429. This exemplifies how we can use the probability notation and formulae to manipulate proportions as well as probabilities.

Instead of working with row totals we can also work with column totals. For example,

$$\text{pr}(\textit{non-smoker} \mid \textit{banned}) = \frac{0.3350}{0.3550} = 0.9437.$$

From this we can say that about 94% of people in the survey who favor banning smoking from the workplace are non-smokers. In answering questions about two-way tables, you will be asked to provide answers in terms of probabilities, proportions or percentages; the general public are often a lot happier with percentages than proportions or probabilities.

Finally we might ask how useful are such proportions. If the 800 respondents represent a random sample of Americans, then we can use these proportions as estimates for the whole population.

Exercises on Section 4.6.1

1. Using Table 4.6.1, what is the probability that a randomly chosen patient has

   (a) nodular cancer, given they have cancer of the head and neck?

   (b) cancer of the head, neck or trunk, given that their cancer is nodular.

*2. In Example 4.5.2 Wild was available with probability 0.7, Seber with probability 0.6 and both with probability 0.5. What is the probability that Seber is available, given that Wild is available?

The conditional probability formula is not difficult to operate. A more difficult skill is recognizing where a conditional probability is required. No one in real life will ever ask you "what is the probability of A given B?" The problem arises in other guises, for example, "if A occurs, how likely is B?", or, in terms of proportions of a population, "what proportion of those with property A also have property B?" In the exercises to follow, we have also included a few unconditional statements to keep you alert.

3. The following relate to Example 4.4.9 and Table 4.4.1.

   (a) **(i)** What is the probability that a random job loss was by a female and the reason given was slack work? **(ii)** Where a job was lost by a female, what is the probability that the reason given was slack work?

   (b) **(i)** What proportion of jobs were lost by males for slack work?
   (ii) What proportion of male job losses were for slack work?
   (iii) What proportion of job losses were due to slack work?

   (c) What proportion of all job losses were due to the position being abolished: **(i)** for females   **(ii)** for males?   **(iii)** in general?

4. The following relate to the data and story of Example 4.5.4 and Table 4.5.1. Parts (a) and (b) concern a randomly selected couple.

   (a) What is the probability that: **(i)** the male is divorced? **(ii)** the female is divorced if the male is divorced? **(iii)** the male partner of a divorced female is divorced? **(iv)** the male is divorced and the female is divorced?

   (b) **(i)** By considering all the relevant conditional probabilities, where the female has been widowed, what is the marital status of the male most likely to be? **(ii)** You can answer part **(i)** without doing the conditional probability calculations. Look at your calculations and see if you can see why (and how) this can be done. **(iii)** If the male is "never married", what is the female most likely to be? What is the probability that she will have this status?

   (c) **(i)** For what proportion of couples in which the male is "never married" is the female also "never married"? **(ii)** For what proportion of couples is the male "never married" and the female "never married"?

   (d) For what proportion of couples in which the male is no longer married (i.e. divorced or widowed) is the female also no longer married?

5. From Table 4.6.2, find the proportion of non-smokers who are in favor of some form of restriction, i.e. having either a ban or special smoking areas.

4.6.2 Multiplication rule

Sometimes the available information is conditional and we have to use it to find the probabilities (or proportions) of other events. This often involves using the so-called multiplication rule (to follow), often as part of a larger calculation.

Example 4.6.4. In 1992, 14% of the population of Israel was Arab, and of those, 52% were described as living below the poverty line. What proportion of the population of Israel consisted of Arabs living below the poverty line?

There are two basic events here "Arab" and living below the poverty line which we shall shorten to "Poor". We see that $\mathrm{pr}(Arab) = 0.14$. The 52%, however, relates only to the Arab subset of the population and is thus conditional. We have $0.52 = \mathrm{pr}(Poor \mid Arab)$. The quantity we want is $\mathrm{pr}(Poor$ and $Arab)$.

(Multiplication Rule)
$$\mathbf{pr}(A \cap B) \; = \; \mathbf{pr}(B)\mathbf{pr}(A|B) \; = \; \mathbf{pr}(A)\mathbf{pr}(B|A)$$

The multiplication rule follows directly from the definition of conditional probability.

Example 4.6.4 cont The multiplication rule gives us

$$\mathrm{pr}(Poor \text{ and } Arab) = \mathrm{pr}(Arab)\ \mathrm{pr}(Poor \mid Arab)$$
$$= \; 0.14 \times 0.52 \; = \; 0.0728,$$

or just over 7% of the population being Arabs living below the poverty line.

A more elementary way of understanding the multiplication rule is as a "proportion of a proportion" or a "percentage of a percentage" – you may recall from basic arithmetic that when you want to find a fraction of a fraction, you multiply. The situation is illustrated in Fig. 4.6.1.
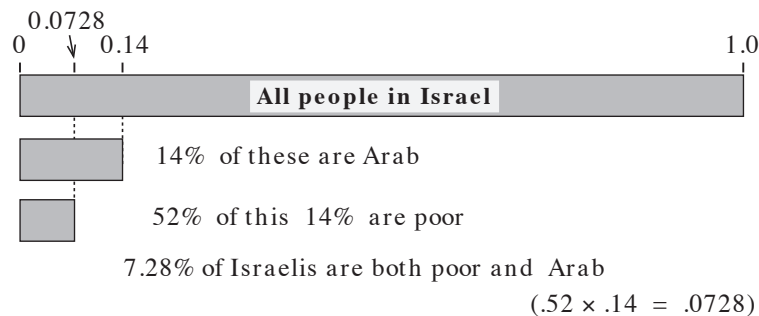

Figure 4.6.1 : Illustration of the multiplication rule.

Remember that a percentage is a shorthand way of expressing a fraction, e.g. $14\% = \frac{14}{100}$. Thus, $14\% \times 52\%$ is really

$$14\% \times 52\% = \frac{14}{100} \times \frac{52}{100} = \frac{728}{10,000} = 7.28\% \,.$$

Example 4.6.5. Two balls are drawn at random, one at a time without replacement, from a box of 4 white and 2 red balls. What is the probability that both balls are white?

Let W_1 and W_2 represent the events "first ball is white" and "second ball is white" respectively, then

$$\mathrm{pr}(W_1 \cap W_2) = \mathrm{pr}(W_1)\,\mathrm{pr}(W_2 \mid W_1).$$

As each ball is equally likely to be selected as the first ball, $\mathrm{pr}(W_1)$ is the number of white balls divided by the number of balls i.e. $\frac{4}{6}$. Once the first ball is chosen, there are 5 balls left of which 3 are white. Hence $\mathrm{pr}(W_2 \mid W_1) = \frac{3}{5}$ and

$$\mathrm{pr}(W_1 \cap W_2) = \frac{4}{6} \times \frac{3}{5} = \frac{2}{5}.$$

Exercises on Section 4.6.2

1. In the US, approximately 1% of the population is schizophrenic, 0.8% of people are homeless, and one third of the homeless are schizophrenic. Using probability notation and the events "*homeless*" and "*schizophrenic*", write down the three pieces of information given here in terms of probability statements about events. (Note: the 1% is inaccurate – see Review Exercises 3, problem 13).

2. According to a survey reported in the *Kitchener-Waterloo Record* (17 May 1989), 26% of residents of the Canadian province of British Columbia aged between 18 and 25 had used cocaine, and 77% of those who tried it once used it again. Based on these figures, what proportion of B.C. residents in this age group had used cocaine at least twice?

3. Referring to Example 4.6.5, what is the probability that both balls are red?

4. 52% of the South Korean workforce works in the service sector of the economy (*TIME*, 2 July 1990). Since 62.5% of Koreans live in South Korea, we shall assume that 62.5% of the workforce lives in the South. What percentage of the entire workforce on the Korean peninsula both lives in the South and works in the service sector?

4.6.3 More complicated calculations using conditional probabilities

Example 4.6.6 Suppose we sample 2 balls at random, one at a time without replacement, from an urn containing 4 black balls and 3 white balls. We want to calculate the probability that the second ball is black. When we come to make the second draw, the chances of drawing a black ball depend upon what ball was removed at the first draw because that determines the composition of the balls in the urn. We will, therefore, have to use information which comes naturally in the form of conditional probabilities. We use the same notation as in Example 4.6.5, e.g. B_2 denotes the event that the second ball sampled is black.

Tree diagrams:

The method we are going to use to tackle problems like the above involves a type of diagram called a ***(probability) tree diagram***. These diagrams often provide a convenient way of organizing (and then using) conditional probability information. To motivate the discussion, Fig. 4.6.2 gives a tree diagram for the situation in Example 4.6.6. Along the way, we will state some general rules for constructing and using such trees.
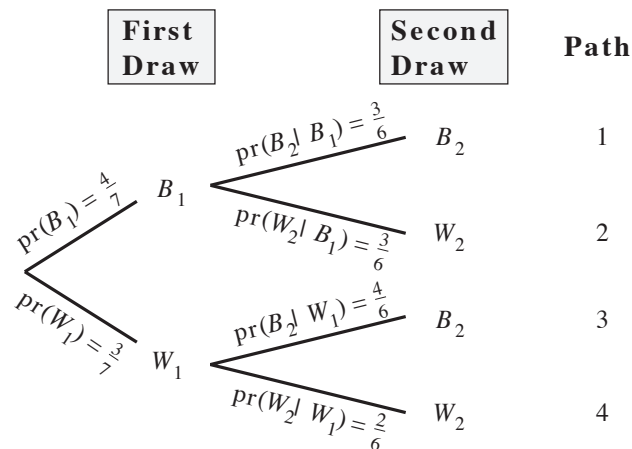
Figure 4.6.2 : Tree diagram for a sampling problem.

The probability written beside each line segment in the tree is the probability that the right hand event on the line segment occurs ***given*** the occurrence of all the events that have appeared along that path so far (reading from left to right). Each time a branching occurs in the tree, we want to cover all eventualities so the probabilities beside any "fan" of line segments should add to unity.

Because the probability information on a line segment is conditional upon what has gone before, the order in which the tree branches should reflect the type of information that is available. In Example 4.6.6, we have unconditional probability information about the first draw so the first set of branches of the tree concern the first draw. The readily available probability information about

the second draw depends upon (i.e. is conditional upon) what happened at the first draw and thus forms the second set of branches. We draw the tree to represent all possible outcomes.

Rules for use

(i) *Multiply along a path* to get the probability that all of the events on that path occur.

(ii) *Add* the probabilities of *all whole paths* in which an event occurs to obtain the probability of that event occurring.

Example 4.6.6 cont. To get the probability of obtaining a black ball on the second draw, the rules tell us to multiply along paths and add whole paths containing a black ball on the second draw (namely, paths 1 and 3). This gives us

$$\mathrm{pr}(B_2) \quad = \quad \frac{4}{7} \times \frac{3}{6} \quad + \quad \frac{3}{7} \times \frac{4}{6} \quad = \quad \frac{4}{7}.$$

Why do these rules work?

Suppose that we have 2 events, A and B, each of which may or may not occur. A tree diagram is presented in Fig. 4.6.3. The order of branching used there is appropriate in situations where the available probability information about B is conditional upon whether or not A occurs.
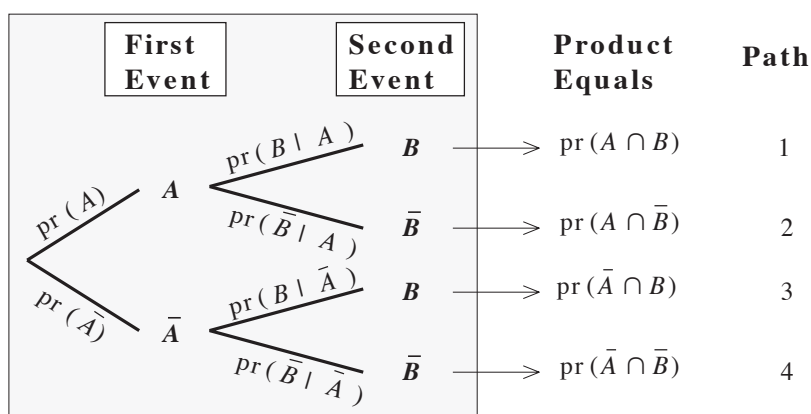

First Event	Second Event	Product Equals	Path
$\mathrm{pr}(B\mid A)$ → B		$\mathrm{pr}(A \cap B)$	1
$\mathrm{pr}(\bar{B}\mid A)$ → $\bar{B}$		$\mathrm{pr}(A \cap \bar{B})$	2
$\mathrm{pr}(B\mid \bar{A})$ → B		$\mathrm{pr}(\bar{A} \cap B)$	3
$\mathrm{pr}(\bar{B}\mid \bar{A})$ → $\bar{B}$		$\mathrm{pr}(\bar{A} \cap \bar{B})$	4

Figure 4.6.3 : Tree diagram for whether or not two events occur.

Applying the Rule (i) to the upper path (Path 1) of Fig. 4.6.3, we get $\mathrm{pr}(A \cap B) = \mathrm{pr}(A)\mathrm{pr}(B\,|\,A)$ as in the third column of the figure. So this first rule governing use of the tree is just a restatement of the multiplication rule.[17] Let us now think about applying the second rule to obtaining $\mathrm{pr}(B)$. This rule follows from the fact that all the paths containing B correspond to mutually

---

[17]Although we have showed this only in a simple case, the reasoning is generally true for any properly laid out probability tree, even if it has multiple branchings. The generalized multiplication rule for chains involving three or more events is given in Section 4.6.4.

exclusive events in which B occurs. Thus, to get $\text{pr}(B)$ we add the probabilities of all such events.

Using the multiplication formula of the previous Section 4.6.2 we can re-express probability Rule 4 (Section 4.5.1) in terms of conditional probabilities, namely:

Alternative expression of Rule*[18] *4

$$\text{pr}(B) = \text{pr}(A \cap B) \ + \ \text{pr}(\overline{A} \cap B) \qquad\qquad \text{(Rule 4)}$$
$$= \text{pr}(A) \, \text{pr}(B \mid A) \ + \ \text{pr}(\overline{A}) \, \text{pr}(B \mid \overline{A}).$$

This is exactly what we get by applying the rules governing the use of the tree (multiply along paths and then add whole paths) to event B.

Example 4.6.7 Of American women using contraception[19]: 38% are sterilized, 32% use oral contraceptives, 24% use barrier methods (diaphragm, condom, cervical caps), 3% use IUD's, and 3% rely on spermicides (foams, creams, jellies). If we define the failure rates of a method as the percentage of who become pregnant during a year of use of the method, then the failure rates for each of these methods are approximately: sterilization 0%, the contraceptive pill 5%, barrier methods 14%, IUD's 6%, and spermicides 26%. What percentage of women using contraception experience an unwanted pregnancy over the course of a year?

In this problem, we have information on the method of contraception used and on failure rates of those contraceptives. Since we are only considering women who are using contraception, the method of contraception forms a partition of the sample space since it divides the population of all such women up into distinct groups. We are given the proportions of women falling into each group, i.e. information of the form $\text{pr}(Method)$. Our information on contraceptive failure depends upon the contraceptive used, i.e. is of the form $\text{pr}(Failure \mid Method)$. Thus the first fan of branches of the tree should be to all of the methods. The second set of branches is then failure or not failure within methods. The tree and all the relevant probabilities are given in Fig. 4.6.4. We have denoted the event of failure[20] by "F". To save space, we have omitted the probabilities in the lower half of most of the pairs of outcomes: the lower probability will be one minus the upper probability.

---

[18]The roles of A and B have been interchanged from the earlier statement of Rule 4 to correspond to the order of events used in the tree diagram.
[19]These figures are from *TIME*, 26 February 1990.
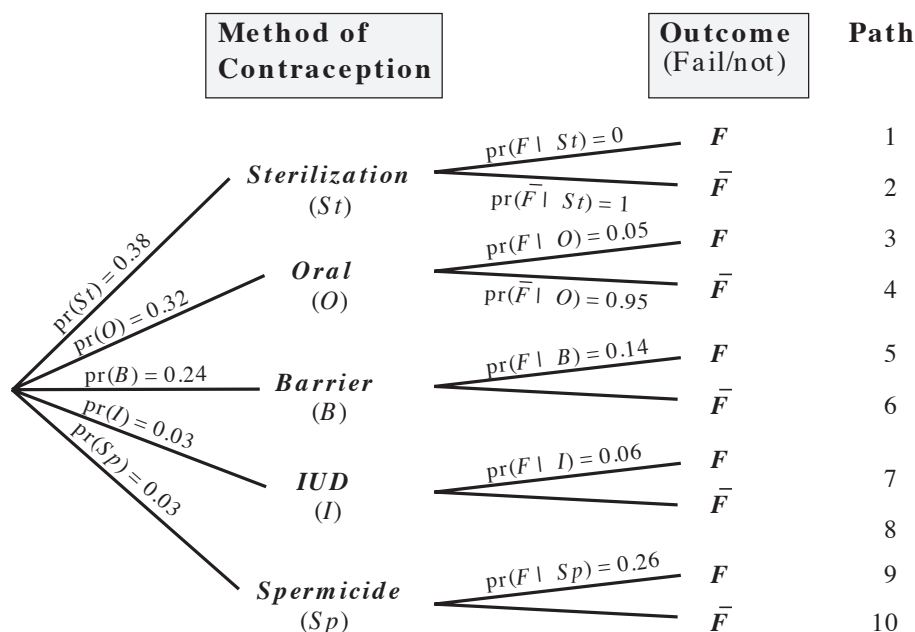[20]In the sense that the woman becomes pregnant during the year.

Figure 4.6.4 : Tree diagram for contraceptive failure.

We can find $\mathrm{pr}(F)$ by applying the two rules governing use of the tree (namely, multiplying along paths and adding whole paths 1, 3, 5, 7, 9 containing F) to obtain

$$\mathrm{pr}(F) = .38 \times 0 \ + \ .32 \times .05 \ + \ .24 \times .14 \ + \ .03 \times .06 \ + \ .03 \times .26$$
$$= 0.0592 \,.$$

Another way of computing this probability by constructing a two-way table is described later in Section 4.8. We can also interpret this probability as a proportion or percentage, namely 5.92% of women in the study experienced contraceptive failure. Finally we note that the usefulness of such proportions or percentages in applying them to the whole population will depend on how the survey was carried out and the size of the survey.

Theory

Let us now discuss the type of situation we saw in Example 4.6.7 in a more general setting. The entire sample space is broken up into separate pieces $C_1, C_2, \ldots C_k$ (i.e. a partition, see Section 4.5.2). We are interested in an event A but our available information is of the form $\mathrm{pr}(C_i)$ and $\mathrm{pr}(A \mid C_i)$. We lay this information out in a tree diagram as in Fig. 4.6.5. Either by applying the rules about trees (multiply along paths and add whole paths) to A, or by applying the multiplication rule to the earlier statement of the Partition Theorem (Section 4.5.2), we obtain the following:

Alternative expression of the Partition Theorem

$$\mathbf{pr}(A) \ = \ \sum_{i=1}^{k} \mathrm{pr}(A \cap C_i) \ = \ \sum_{i=1}^{k} \mathbf{pr}(C_i)\mathbf{pr}(A|C_i).$$

In our experience, many people find it easier to solve problems by using trees to organize their thinking pictorially than by using formulae, particularly when they are just beginning to learn to manipulate conditional probabilities. However, it is essential that we are dealing with a partition which we can check by seeing if the probabilities labeling the first fan of branches add to unity.
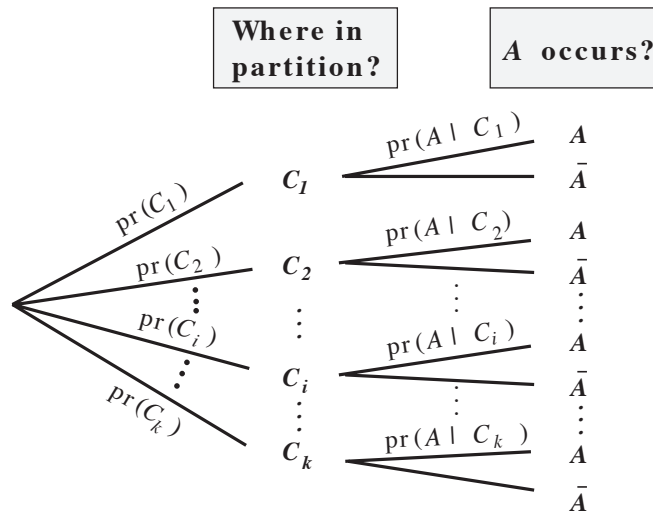

Figure 4.6.5 :   Tree diagram for the partition problem.

Two Case Studies showing the application of these ideas to real problems follow before the Exercises on this subsection. Both Case Studies illustrate use of the situation depicted in Fig. 4.6.3.

*Case Study 4.6.1   Randomized response

Imagine, as part of a survey, asking the question, "Have you ever physically or sexually abused your children?" Clearly most child abusers are not going to admit to it and say "Yes" in a face to face interview. Such a survey will grossly under-report the prevalence of child abuse in the population under study. If the personal element is reduced, e.g. by using a telephone rather than a face to face interview, or better still a self-administered questionnaire, the under-reporting will be decreased. However, it will still be present so long as respondents think that someone/anyone will know what answer they gave. Such under-reporting can be expected for any question in which a "Yes" answer is an admission of doing or thinking something which tends to be thought of as socially undesirable e.g. drunk driving, some form of criminal activity, not voting. Randomized response is a strategy for obtaining an estimate of the proportion of people who would answer "Yes" to a threatening question without the interviewer knowing whether a particular respondent is admitting to the activity or not. There are many slight variants but here is one version.

Instead of just presenting the (threatening) question of interest, a pair of questions are presented, the real question printed in red and a nonthreatening dummy question with a known response rate, printed in green. We shall use,

"Was your mother born in May?" as our dummy. The proportion of people born in May should be obtained from birth statistics, but we shall approximate it by $\frac{1}{12}$. A random mechanism is then used to choose which question the respondent is to answer. Locander et al. [1976] use a box with 50 colored beads, 70% of which are red and 30% are green. The box is shaken and one bead appears at a small window. If a red bead shows, the respondent answers the red question. If the green bead shows, they answer the green question. The interviewer does not see the bead and thus does not know which question is being answered. The situation is depicted in Fig. 4.6.6.
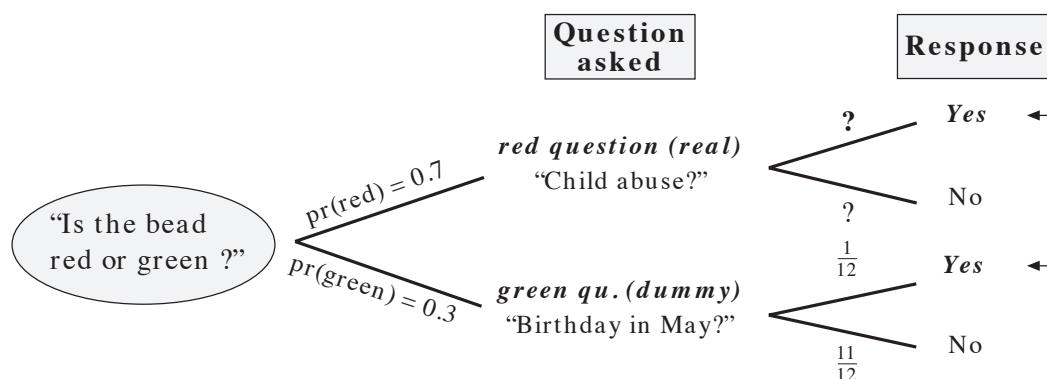

Figure 4.6.6 : Randomized response.

Suppose 8% (0.08) of our sample answered "Yes". Now using the tree (multiply along paths and add whole paths containing "Yes"), we obtain

$$\text{pr}(\textit{answering yes}) = \text{pr}(\textit{getting red}) \ \text{pr}(\textit{answering yes} \mid \textit{red})$$
$$+ \ \text{pr}(\textit{getting green}) \ \text{pr}(\textit{answering yes} \mid \textit{green}).$$

Using the sample proportion of 0.08 to approximate pr(*answering yes*), we obtain

$$0.08 \ = \ 0.7 \times \text{pr}(\textit{answering yes} \mid \textit{child-abuse question}) \ + \ 0.3 \times \tfrac{1}{12}.$$

We solve this equation to obtain the probability of answering yes to the child-abuse question, namely

$$\text{pr}(\textit{answering yes} \mid \textit{child-abuse question}) \ = \ \frac{0.08 - 0.3 \times \frac{1}{12}}{0.7} \ = \ 0.79.$$

For honest replies it is important that the respondent is convinced that the interviewer does not know which question he or she is answering. Thus it must be clear that the interviewer cannot know the correct answer to the dummy question and "Yes" answers to the dummy cannot be so rare that the respondent gets suspicious. The technique is reasonably effective but not perfect. For example, Locander et al. [1976] experienced some under-reporting in a conviction for drunken driving question in which the responses could be

validated against conviction records. Still, randomized response resulted in considerably less under-reporting than any of the interview methods (face-to-face, telephone or self-administered). Only 1 in 20 respondents found the randomized response technique confusing, silly or unnecessary.

Case Study 4.6.2   How big a problem is AIDS in your community?

It is well known that AIDS is one of the most important public health problems facing the world today. AIDS is believed to be caused by the human immunodeficiency virus (HIV), but many years can elapse between HIV infection and the development of AIDS. In 1990, the World Health Organization (WHO) projected between 25 and 30 million cases of HIV infection world wide by the year 2000. The United States, where 200,000 AIDS cases had been reported by mid 1992, has been the worst affected western country, largely because the epidemic began earlier in the US In 1990, WHO estimated that one in every 75 males and one in every 700 females in the US was infected with HIV. Numbers of AIDS cases in some English speaking countries are given in Table 4.8.2.

The enzyme-linked immunosorbent assay (ELISA) test is used to screen blood samples for antibodies to the HIV virus (rather than the virus itself). It gives a measured "mean absorbance ratio" for HIV (previously called HTLV) antibodies. Table 4.6.3 gives the absorbance ratio values for 297 healthy blood donors and 88 HIV patients. Healthy donors tend to give low ratios but some are quite high, partly because the test also responds to some other types of antibody e.g. human leucocyte antigen or HLA (Gastwirth [1987, page 220]). HIV patients tend to have high ratios but a few give lower values because they have not been able to mount a strong immune reaction. To use this test in practice, we need a cutoff value so that those who fall below the value are deemed to have tested negatively and those above to have tested positively. Any such cutoff will involve misclassifying some people without HIV as having a positive HIV test (which will be a huge emotional shock), and some people with HIV as having a negative HIV test (with consequences to their own health, the health of people about them, the integrity of the blood bank, ...). Using a cutoff ratio of 3 we find that of the healthy people[21] (Table 4.6.3), $275/297 = 0.926$ test negatively (22 false positives) and for HIV patients $86/88 = 0.98$ test positively (2 false negatives). It should be noted that the false negative rate may be an undercount.[22] Better results than these have been obtained with the multiple use of ELISA (Gastwirth, 1987 page 236) and with modern commercial versions of the test. The proportions given above are only rough estimates from small

---

[21]In the medical and biostatistical literatures, the probability of correctly diagnosing a "sick" individual as "sick" is called the *sensitivity* of a test, while the probability of correctly diagnosing that a "healthy" individual does not have the condition of interest is called the *specificity* of that test.

[22]It appears that the virus takes 6 to 12 weeks to provoke antibody production (*TIME*, 2 March 1987, page 44). Also, *TIME* (12 June 1989) reports cases of infected men who had not produced antibodies for up to 3 years.

samples. Nevertheless, in what follows we shall use them as if they were true probabilities.

Hence, $\qquad\qquad\qquad\qquad$ $\mathrm{pr}(Positive \mid HIV) = 0.98.$

Since for people without HIV, the test is negative with probability 0.926, it is positive for these people with probability $1 - 0.926 = 0.074$. We shall round this value (as the information is very approximate) and use

$$\mathrm{pr}(Positive \mid No\ HIV) = 0.07.$$

Table 4.6.3 : Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies[a]

MAR	Healthy Donor	HIV patients
< 2	202 ⎫ 275	0 ⎫ 2
2 − 2.99	73 ⎭	2 ⎭
3 − 3.99	15	7
4 − 4.99	3	7
5 − 5.99	2	15
6 − 11.99	2	36
12+	0	21
Total	297	88

[a]From Gastwirth [1987, Table 4].

Suppose now you wish to estimate the proportion, $\mathrm{pr}(HIV)$, of people in your community with HIV. The appropriate probability tree is given in Fig. 4.6.7. Unfortunately, to estimate the proportion with HIV, you cannot just sample people and use the proportion testing positively as an estimate. Suppose, for the sake of exposition, that 1% of people have HIV. In the discussion which follows, we make use of the numerical equivalence between proportions of a population and probabilities for a randomly chosen individual (Section 4.4.5). From the tree, we have

$$\mathrm{pr}(Positive) \;=\; \mathrm{pr}(HIV) \times 0.98 \;+\; (1 - \mathrm{pr}(HIV)) \times 0.07. \qquad (1)$$

Suppose, for example, that 1% of people have HIV, i.e. $\mathrm{pr}(HIV) = 0.01$. Then, equation (1) gives $\mathrm{pr}(Positive) = 0.079$, which tells us that approximately 8% will test positively. Under these circumstances, the large majority of the people in any sample who test positively will not in fact have HIV! They are so-called *"false positives"*.
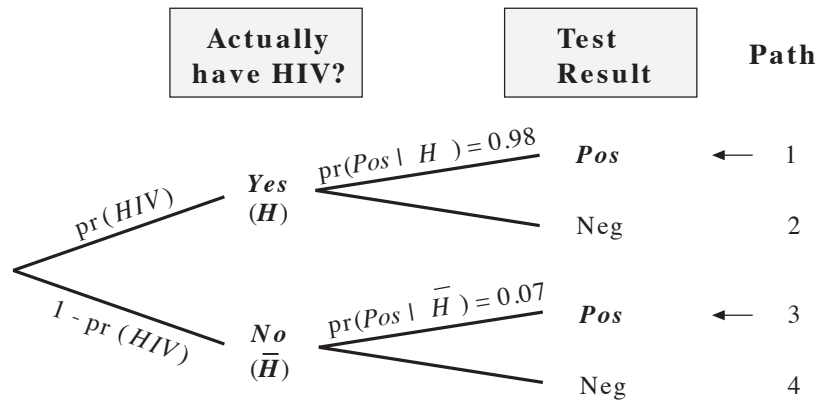
Figure 4.6.7 :   Probability tree for HIV testing.

In reality, $\mathrm{pr}(HIV)$, the proportion of people with HIV, will be unknown. It turns out that we can use equation (1) above to obtain a good estimate of the unknown value of $\mathrm{pr}(HIV)$ as follows. First, we can take a sample to get a good estimate of the proportion of people in the community who would test positively. We then replace $\mathrm{pr}(Positive)$ in (1) by its sample estimate and solve[23] for $\mathrm{pr}(HIV)$. Suppose that 9% of your sample tests positively, then

$$0.09 \simeq \mathrm{pr}(HIV) \times 0.98 + (1 - \mathrm{pr}(HIV)) \times 0.07,$$

which gives

$$\mathrm{pr}(HIV) \simeq \frac{0.09 - 0.07}{0.98 - 0.07} = 0.022.$$

Thus if 9% of the sample tested positively we would estimate that only 2% of the population was actually infected by HIV.[24]

Exercises on Section 4.6.3

1. Referring to Example 4.6.6, what is the probability that
   (**a**) both balls are white,
   (**b**) the first is red and the second is white, and
   (**c**) they are of different colors?

2. In NZ, 3.24% of Europeans and 1.77% of Maori have type AB blood. A blood bank in a district where the population is 85% European and 15% Maori wants to know how much AB blood to stock. What percentage of people in the district have AB blood?

3. 27% of the North Korean workforce works in the service sector of the economy compared with 52% of the South Korean workforce (*TIME*, 2 July

---

[23]Gastwirth [1987] discusses the statistical errors associated with such estimates.

[24]The same methodology can be used with other imperfect test procedures, e.g. to estimate the proportion of people who are lying using the lie detector test, the proportion of job applicants who actually have sufficient skills to do the job given they are selected on the basis of their scores on a test given to all applicants, the proportions of people on drugs from drug screening programs.

1990). Now 62.5% of the population of the Korean peninsula lives in South Korea; the rest live in the North. We shall assume that the workforce of Korea is divided in the same proportions. What percentage of the entire workforce on the Korean peninsula works in the service sector?

4. The chances of a child being left-handed are 1 in 2 if both parents are left-handed, 1 in 6 if one parent is left-handed and 1 in 16 if neither parent is left-handed (*NZ Herald*, 5 January 1991). Suppose that of couples having children, in 2% both father and mother are left-handed, in 20% one is a left-handed and in the rest, neither is a left-handed. What is the probability of a randomly chosen child being a left-handed?

*4.6.4 Generalized multiplication rule[25]

The multiplication rule can be extended to more than two events. For example, applying the rule successively to the pairs of events $A_1 \cap A_2$ and A_3, and then A_1 and A_2, gives us

$$\mathrm{pr}(A_1 \cap A_2 \cap A_3) = \mathrm{pr}(A_1 \cap A_2)\,\mathrm{pr}(A_3 \mid A_1 \cap A_2)$$
$$= \mathrm{pr}(A_1)\,\mathrm{pr}(A_2 \mid A_1)\,\mathrm{pr}(A_3 \mid A_1 \cap A_2).$$

This rule is useful for finding the probability of the joint occurrence of A_1, A_2 and A_3 when we are given $\mathrm{pr}(A_1)$ and the conditional probabilities $\mathrm{pr}(A_2 \mid A_1)$ and $\mathrm{pr}(A_3 \mid A_1 \cap A_2)$. The latter probability is the probability that A_3 occurs, given that *both* A_1 and A_2 have occurred. What events are labeled A_1, A_2, or A_3 (and thus the order in which the conditioning is used) is arbitrary. In practice the order used is determined by the sort of conditional probability information you have. Quite often, the order used is the order that the events occur in time (as we see by the following example).

Example 4.6.8. Three balls are drawn without replacement from a box with w white balls and r red balls. What is the probability of getting the sequence white, red, white? The answer is given by

$$\mathrm{pr}(W_1 \cap R_2 \cap W_3) = \mathrm{pr}(W_1)\,\mathrm{pr}(R_2 \mid W_1)\,\mathrm{pr}(W_3 \mid W_1 \cap R_2)$$
$$= \frac{w}{w+r} \cdot \frac{r}{w+r-1} \cdot \frac{w-1}{w+r-2}.$$

Tree diagrams can again be used to solve this type of problem involving a chain of events. We simply add further branches to a figure like Fig. 4.6.3.[26]

---

[25]This topic, which is included for completeness only, is not required elsewhere in this book.
[26]Note that if the conditional probabilities are laid out on a tree following the usual convention by which the probability written by a line segment is the conditional probability of the event following given the occurrence of all the events that have appeared before, the generalized multiplication rule is equivalent to rule (i) for trees, namely, the probability that all events represented along a path occur is obtained by multiplying all the probabilities along that path.

The formula for the extension of the multiplication rule to ***an arbitrary number of events*** follows the pattern above, namely of introducing a new event each time and conditioning upon every event that has been used before. Thus

$$\mathrm{pr}(A_1 \cap A_2 \cap \ldots A_n) = \mathrm{pr}(A_1)\,\mathrm{pr}(A_2 \mid A_1)\,\mathrm{pr}(A_3 \mid A_1 \cap A_2)\ldots$$
$$\mathrm{pr}(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Quiz on Section 4.6

1. In $\mathrm{pr}(A \mid B)$, how should the symbol "$|$" be read? (Section 4.6.1)
2. Give an example where A and B are two events with $\mathrm{pr}(A) \neq 0$ but $\mathrm{pr}(A \mid B) = 0$. (Section 4.6.1)
3. If event A always occurs when B occurs, what can you say about $\mathrm{pr}(A \mid B)$? (Section 4.6.1)
4. When drawing a probability tree for a particular problem, how do you know what events to use for the first fan of branches and which to use for the subsequent fans? What probability do you use to label a line segment?
   How do you find the probability that all events along a given branch occur?
   How do you find the probability that a particular event occurs? (Section 4.6.3)
5. Describe in words the generalized multiplication rule. (Section 4.6.4)

4.7 Statistical Independence

4.7.1 Two events

We have seen (e.g. Example 4.6.3) that our assessment of the chances that an event occurs can change drastically depending upon the information we have about other events. In general, $\mathrm{pr}(A \mid B)$ and $\mathrm{pr}(A)$ are different. However, if there is no change, i.e. $\mathrm{pr}(A \mid B) = \mathrm{pr}(A)$, then knowing whether B has occurred gives no new information about the chances of A occurring. We then say that A and B are ***statistically independent*** and we have the following definition:

> Events A and B are statistically independent if
> $$\mathrm{pr}(A \mid B) = \mathrm{pr}(A).$$

In this case, the multiplication formula $\mathrm{pr}(A \cap B) = \mathrm{pr}(B)\,\mathrm{pr}(A \mid B)$ becomes $\mathrm{pr}(A \cap B) = \mathrm{pr}(A)\,\mathrm{pr}(B)$ and we take this as our working rule.

> If A and B are statistically independent, then
> $$\mathrm{pr}(A \cap B) = \mathrm{pr}(A)\,\mathrm{pr}(B).$$

From the working rule we find that we also have $\mathrm{pr}(B \mid A) = \mathrm{pr}(B)$ so that it doesn't matter whether our definition of independence is given in terms of

$\mathrm{pr}(A \mid B)$ or $\mathrm{pr}(B \mid A)$: if A is independent of B, then B is also independent of A. Our main use of the working rule is to calculate $\mathrm{pr}(A \cap B)$ when we know that A and B are independent. It can also be used to check for statistical independence in a probability model by checking whether the formula works (see Example 4.7.2 to follow).

Unfortunately, it is only possible to establish statistical independence theoretically in rather trivial models. To establish that events are exactly independent in a more complex setting by doing an experiment and collecting data would require an infinite amount of data. In most statistical modeling of the real world, independence is assumed when it seems reasonable to do so on the basis of subject-matter knowledge or intuition. The notion of *physical independence* gives one way of thinking about this important modeling issue. Events are *physically independent* if there is no physical way in which the outcome of one event can influence the outcome of the other. Under any sensible probability model, physically independent events will be statistically independent.

Note: If A and B are independent, then so are[27] A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$. This makes intuitive sense for physically independent experiments. Suppose A can occur in experiment 1 and B in experiment 2. If the two experiments are physically independent, then the occurrence or non-occurrence of A will not affect the occurrence or non-occurrence of B.

Example 4.7.1   According to one study, 30% of Black Americans have type A blood and 26% have the genetic marker PGM 1–2. Now if these characteristics appear independently, and there is evidence that they do, then the probability that a Black American has both is

$$\mathrm{pr}(\textit{Type A} \text{ and } \textit{PGM 1–2}) = \mathrm{pr}(\textit{Type A}) \times \mathrm{pr}(\textit{PGM 1–2}) \qquad \text{(as indep.)}$$
$$= 0.3 \times 0.26 = 0.078.$$

The probability that a black American has Type A blood but does not have PGM 1–2 is

$$\mathrm{pr}(\textit{Type A} \text{ and } \overline{\textit{PGM 1–2}}) = \mathrm{pr}(\textit{Type A}) \times \mathrm{pr}(\overline{\textit{PGM 1–2}}) \qquad \text{(as indep.)}$$
$$= 0.3 \times 0.74 = 0.222.$$

Example 4.7.2   Suppose we toss a coin twice, so that $S = \{HH, HT, TH, TT\}$. Let

$$A = \text{“at least one head”} = \{HH, HT, TH\}$$
$$B = \text{“at least one tail”} = \{HT, TH, TT\}.$$

Assuming equally-likely outcomes, we have $\mathrm{pr}(A) = \frac{3}{4}$ and $\mathrm{pr}(B) = \frac{3}{4}$. However, the probability that both events occur is

$$\mathrm{pr}(A \cap B) = \mathrm{pr}(\{HT, TH\}) = \tfrac{1}{2} \neq \mathrm{pr}(A) \times \mathrm{pr}(B).$$

---

[27]These can be proved formally using the probability rules.

Thus the events A and B are *not* independent. However, let

$$C = \text{“head at first toss”} = \{HH, HT\}$$
$$D = \text{“head at second toss”} = \{HH, TH\}.$$

Now $\text{pr}(C) = \text{pr}(D) = \frac{2}{4} = \frac{1}{2}$ and

$$\text{pr}(C \cap D) = \text{pr}(\{HH\}) = \frac{1}{4} = \text{pr}(C) \times \text{pr}(D).$$

Thus, under the assumption of equally likely outcomes, these two events *are* independent, as we would expect from physical intuition. It is reassuring that the model assuming equally-likely outcomes is providing sensible answers.

Example 4.7.3 We look again at Example 4.6.5 in which two balls are drawn one at a time from a box of 4 white and 2 red balls. Suppose this time that the balls are drawn at random *with* replacement. Because the first ball is returned before the second draw is made, the outcome of the first draw no longer changes the composition of balls in the urn and thus has no influence on the outcome of the second draw, i.e. they are physically independent.[28] Because the events are physically independent, they are also statistically independent implying that the probabilities on the second draw are not affected by what happened on the first draw, e.g. $\text{pr}(W_2 \mid W_1) = \text{pr}(W_2 \mid R_1) = \text{pr}(W_2)$. You can check that these probabilities are in fact all identical here and equal here to $\frac{4}{6}$ by reasoning from the physical situation (note that $\text{pr}(W_1) = \frac{4}{6}$ as well).

■

We can still use ***tree diagrams*** with independent events. In fact they are simpler to work with because the conditional probabilities are the same as unconditional probabilities under independence. Thus, we can just use unconditional probabilities to label the tree. The rules for using the tree are unchanged.

Example 4.7.3 cont. This is illustrated in Fig 4.7.1 which depicts the situation in Example 4.7.3. For this example, as the events are independent,

$$\text{pr}(W_1 \cap R_2) = \text{pr}(W_1)\,\text{pr}(R_2) = \frac{4}{6} \times \frac{2}{6} = \frac{8}{36}.$$

The probability of obtaining one ball of either color is

$$\text{pr}(W_1)\text{pr}(R_2) + \text{pr}(R_1)\text{pr}(W_2) = \frac{4}{6} \cdot \frac{2}{6} + \frac{2}{6} \cdot \frac{4}{6} = \frac{16}{36}.$$

---

[28]We use the same notation as in Example 4.6.6, i.e. W_1 and W_2 represent the events “first ball is white” and “second ball is white” respectively, etc.
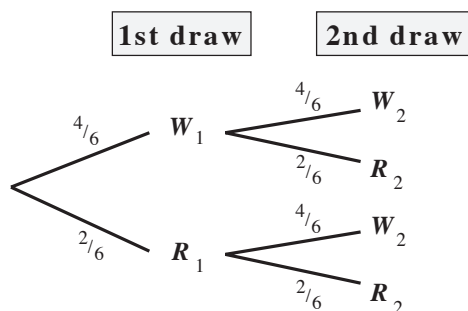
Figure 4.7.1 : Probability tree for Example 4.7.3.

Exercises on Section 4.7.1

1. According to a study on $3,433$ women conducted by the Alan Guttmacher Institute in the US (*Globe and Mail*, 7 August 1989), 6% of women on the contraceptive pill can expect to become pregnant in the first year compared with 14% of women who do not use the pill but whose partners use condoms. What are the chances of the woman becoming pregnant in the first year if she is on the pill *and* he uses condoms? (Assume independence. Is this a reasonable assumption?)

2. According to a poll conducted for *TIME* (29 January 1991) 42% of American gun owners keep their guns in their bedrooms and 12% keep them loaded all of the time. Assuming independence, what percentage of gun owners keep a loaded gun in their bedroom?

3. White North Americans in California have blood phenotypes A, B, O and AB with probabilities 0.41, 0.11, 0.45 and 0.03 respectively. If two Whites are chosen at random, what is the probability that they have the same phenotypes? Why may you assume that the two are independent?

4. Consider the sexes of children in a 3-child family so that $S = \{GGG, GGB, GBG, BGG, GBB, BGB, BBG, BBB\}$. Assume equally likely outcomes and imitate the working of Example 4.7.2 to show that:

(a) the events $A = $ "eldest child is a girl" and $B = $ "at least one child of each sex" are independent; and

(b) the events B above and $C = $ "at least 2 girls" are also independent.

(c) Consider now a 4-child family. Let the events B and C be described by the same verbal phrases as in (a) and (b). Show that this time B and C are *not* independent!

4.7.2 Positive and negative association

In humans, independence of characteristics, as in Example 4.7.1, tends to be the exception rather than the rule. Some things we know tend to go together, for example blond hair and blue eyes. Someone with blond hair is much more likely to have blue eyes than someone with brown or black hair. We say the events "having blond hair" and "having blue eyes" are positively associated.

Suppose we look at the population in which 30% have blond hair and 25% have blue eyes. If we assumed independence, we would say that the proportion with both is

$$\text{pr}(blond\ hair \cap blue\ eyes) = \text{pr}(blond\ hair)\,\text{pr}(blue\ eyes) = 0.3 \times 0.25 = 0.075.$$

Since these events are not independent, we should have used

$$\text{pr}(blond\ hair \cap blue\ eyes) = \text{pr}(blond\ hair)\,\text{pr}(blue\ eyes \mid blond\ hair)$$
$$= 0.3 \times \textbf{???}.$$

Now among blond-haired people, the proportion with blue eyes is high, probably much closer to 80% than 25%. The product 0.3×0.8 is then much larger than 0.075. Assuming independence when events are positively associated can lead to answers that are far too small.

Many human characteristics are negatively associated as well, i.e. if you have one you are much less likely to have the other. Black hair and blue eyes is one example. Assuming independence when events are negatively associated leads to answers which are too big.

*4.7.3 Mutual independence of more than two events

The n events $A_1, A_2, \ldots A_n$ are ***mutually independent*** if

$$\text{pr}(A_1 \cap A_2 \cap \ldots \cap A_n) = \text{pr}(A_1)\,\text{pr}(A_2)\ldots\text{pr}(A_n),$$

and the same type of multiplication formula holds for *any* subcollection of the events.[29] We shall use this result often in Chapter 5 for events like tosses of a coin which we know are physically independent. However, the formula above and the independence assumption are frequently abused. Often it is relatively easy to get information about individual probabilities, e.g. the proportions of individuals who own their own houses, who believe in abortion, who have a high intelligence, and who hold strong religious beliefs. To calculate the probability that all these conditions hold at the same time we need the multiplication rule of Section 4.6.4. This requires information about conditional probabilities, e.g. the proportion of strongly religious people *among those* who have high intelligence and believe in abortion and own their own house etc. Clearly information on society broken down to this level is hard to find. What often happens is that, in the *absence of knowledge* of the appropriate conditional probabilities, people *assume* independence. From the discussion of the previous section, this can lead to answers that are grossly too small or grossly too large - and we won't know! The two case studies which follow the Exercises for this section contain salutary tales.

---

[29]This is stronger than just requiring every pair of events to be independent; i.e. $\text{pr}(A_i \cap A_j) = \text{pr}(A_i)\,\text{pr}(A_j)$ all $i \neq j$. If $n = 4$, the product rule has to hold not only for all four events but also for any three events and for any two events.

Exercises on Section 4.7.3

1. In April 1990, Mary Ayala was due to give birth to a baby girl conceived to serve as a bone marrow donor for her 17 year-old sister Anissa who had a virulent form of leukemia. Our information came from a *TIME* magazine article (5 March 1990, page 41) focusing principally on ethical considerations involved in the Ayala's actions. However, the article also discussed some probability calculations which showed that when they started out on this course of action, the Ayalas had a slim chance of success. Abe Ayala (the husband) had a vasectomy 16 years before. The chances of successfully reversing such a vasectomy were put at 50%. Mary Ayala was 42 and the chances of a woman aged 40 to 44 conceiving were stated to be 73%. The chance of siblings of the same parents having matched blood marrow were quoted as 25% and the chance of a bone-marrow transplant curing leukemia in this patient was said to be 70%. To be successful, all of these things had to happen.[30]

   (a) Assuming independence, what were the Alayas' chances of success?

   (b) Which criteria do you think are independent and which are you doubtful about?

2. In 1991, a power failure at an AT&T switching center triggered loss of telephone service to over 1 million people in New York City and caused havoc especially with air traffic (*TIME*, 30 September, 1991). When AT&T switched to their own power generation equipment, a power surge tripped a battery powered emergency backup system. This triggered battery-powered alarms that the backup system had been activated but the alarm systems audio-sirens did not come on and the visual warnings were not noticed for over five hours hours. When the batteries ran down the resulting power failure immediately shut off three huge switches that route telephone calls. Which of the events "power failure", "sirens not working", "visual signals not noticed", "batteries ran down" and "routing switches shut off" do you think are likely to be independent? Which seem to be direct causes and effects? Which are likely to be related but do not appear to be causes and effects?

Case Study 4.7.1 People versus Collins

One of the first occasions in which a conviction was obtained in an American court largely on statistical evidence was the case of "People versus Collins". In 1964, Mrs Juanita Brooks was knocked over while walking home with her shopping basket. When she got up she saw a young woman running away and found that her purse was missing. The young woman was described as having blond hair in a pony tail, and as wearing something dark. Another witness, one John Bass, saw such a woman get into a yellow car driven by a black male with a beard and mustache. Collins and his wife Janet fitted the description.

---

[30] *TIME*, 23 March 1992 reported a happy ending. The transplant seemed to be a success. There was no sign of the disease and Anissa was getting married.

Bass picked out Collins in a line-up but there were problems with the identification. To help what may have been a weak identification, the prosecutor called on a college mathematics instructor. This witness explained the product rule above for probabilities of mutually independent events. The prosecutor continued by having the mathematical witness apply the product rule to this case, which he proceeded to do. He assumed the following probabilities (really relative frequencies here) given in Table 4.7.1 for each of the characteristics. Using the product rule to obtain the chances that a random couple meets all the characteristics in the description above, he multiplied the individual probabilities to obtain $\frac{1}{10} \times \frac{1}{4} \times \ldots \times \frac{1}{1000} = 1$ chance in 12 million. The chances of finding such a couple was so overwhelmingly small that the possibility of the police finding another couple fitting the description probably never entered the jurors' heads. The jury was convinced and the Collins couple were convicted.

Table 4.7.1 : Frequencies Assumed by the Prosecution

Yellow car	$\frac{1}{10}$	Girl with blond hair	$\frac{1}{3}$
Man with mustache	$\frac{1}{4}$	Black man with beard	$\frac{1}{10}$
Girl with ponytail	$\frac{1}{10}$	Interracial couple in car	$\frac{1}{1000}$

In 1968 the Californian Supreme Court threw the verdict out. Some of the holes in the argument should be clear from our earlier discussions. Some of the characteristics above are clearly not independent, e.g. man with a mustache and black man with a beard, since most men with beards also have mustaches. Furthermore, if you have a "girl with blond hair" and "negro man with beard" the chance of having an "inter-racial couple" is close to 1 not $\frac{1}{1000}$, so that from this alone the answer is too small by a factor of about 1000. Also, the prosecution had presented no evidence to support the values chosen for their probabilities.

The defense also presented a much more subtle probability argument. The police found one couple fitting the description so at least one such couple existed. The defense calculated the conditional probability that two or more such couples exist given that at least one couple exists. This probability turns out to be is quite large (about 35%), even using the prosecution's 1 in 12 million figure. Thus reasonable doubt that the Collins committed the crime has been created.[31]

Case Study 4.7.2   Nuclear reactor safety

Speed [1977] reviewed the use of probability arguments in the Reactor Safety Study, a major US government study of the safety of nuclear power (see US Government, 1975). This study had come out with estimates like one chance in

---

[31] For further information about this case, see Jonakait [1983], Fairley and Mosteller [1974].

20, 000 per reactor per year of a core meltdown; one chance in a thousand million per reactor per year that containment would fail, releasing virtually all the volatile and gaseous fission products into the atmosphere; one chance in sixteen million per year of an individual in the US being killed by a reactor accident; and so on. How reliable were these figures? We won't go into detailed technical details but will focus upon some elementary issues.

The calculations were based upon a *fault tree analysis*, which is a diagram linking all the things that might go wrong and cause disaster. Hundreds of probability statements were made about things like the following:

  (i) The pump will work when required.
 (ii) The operator will turn on the switch when required.
(iii) The safety system is undergoing maintenance when required.
(iv) Under stated conditions a steam explosion will occur.

Among more technical criticisms, Speed [1977] condemned the Reactor Safety Study on three elementary grounds:

(a) Individual probabilities ascribed to events were based upon little or no data and were sometimes purely subjective.

(b) There were unfounded assumptions of independence in chains of events which could cause gross underestimates of the probability.

(c) Fault tree analysis can only consider the chances of failure from an *anticipated* cause. It is possible that an unanticipated cause of failure may have a reasonably large probability of occurring.

While the study was in Draft form, there was an incident in which the Browns Ferry plant in Alabama was closed down due to a large electrical fire in the control room. The Draft had not even considered such contingencies and had to be modified to include a statement to this effect. The same incident helps illustrate (b). Reasoning of the following form was used.

$$
\begin{array}{c}
\boxed{\begin{array}{c}\textbf{\textit{Probability}}\\\textbf{\textit{of worst}}\\\textbf{\textit{consequence}}\end{array}}
=
\boxed{\begin{array}{c}\textbf{\textit{Probability}}\\\textbf{\textit{of initiating}}\\\textbf{\textit{event}}\end{array}}
\times
\boxed{\begin{array}{c}\textbf{\textit{Probability}}\\\textbf{\textit{of safety}}\\\textbf{\textit{system}}\\\textbf{\textit{failure}}\end{array}}
\times
\boxed{\begin{array}{c}\textbf{\textit{Probability}}\\\textbf{\textit{of worst}}\\\textbf{\textit{weather}}\end{array}}
\times
\boxed{\begin{array}{c}\textbf{\textit{Probability}}\\\textbf{\textit{of highest}}\\\textbf{\textit{population}}\\\textbf{\textit{density}}\end{array}}
\end{array}
$$

Note the assumed independence of "initiating event" and "safety system failure". However, rather than being independent, some initiating events such as large scale electrical fires can actually contribute to safety system failure. Other assumptions were made that people either noticed or did not notice warning signs on a meter *independently*. Yet sometimes an event that causes one person not to notice the dial, e.g. a smoke filled room and the panic of a fire, will cause others not to notice it either. There are therefore positive associations between events in both of these examples, and, as we know from Section 4.7.2, assuming independence when there are positive associations can cause gross underestimates of probabilities. And Speed's conclusions from all this? "...

it would be nothing short of a miracle if the overall system probabilities ever bore any relation to reactor experience."[32]

The moral of this subsection is that we should be highly suspicious of any probabilities obtained under independence assumptions unless convincing reasons or data are given to justify the independence.

Quiz on Section 4.7

1. What does it mean for two events A and B to be statistically independent?
2. Why is the working rule under independence, i.e. $\mathrm{pr}(A \cap B) = \mathrm{pr}(A)\mathrm{pr}(B)$, just a special case of the multiplication rule of Section 4.6.2? What two uses do we make of the working rule? (Section 4.7.1)
3. What do we mean when we say two human characteristics are positively associated? Negatively associated? (Section 4.7.2)
4. What happens to the calculated $\mathrm{pr}(A \cap B)$ if we treat positively associated events as independent? If we treat negatively associated events as independent? (Section 4.7.2)
5. Why do people often treat events as independent? When can we trust their answers? (Section 4.7.3)
6. What is an inherent difficulty with a fault tree for calculating risks? (Case Study 4.7.2)

4.8   Reversing the Order of Conditional Probabilities

Example 4.8.1   Case Study 4.6.2 contained data on the performance of a version of the ELISA test used to try to detect HIV[33] infection in blood. Suppose that all residents of a large city are tested and that 1% of people in that city are actually infected with HIV. On the basis of the information given in Case Study 4.6.2, approximately 98% people who are infected with HIV test positive for HIV, while approximately 93% of people who are uninfected test negative.

What is the probability that a randomly chosen person has HIV given that he or she tested positive? It must be pretty high, right? After all, even though the test is not perfect, it almost always gives the correct answers. We shall find that any such intuition has misled us.

The information we have is repeated in Fig. 4.8.1. We have unconditional information about the probability the person has HIV without reference to the test (1%). Our information about the test's performance is conditional on whether or not the person has HIV. What is new about this example is that in the probability we want, $\mathrm{pr}(HIV \mid Positive)$, the *order of the conditioning is reversed* from that in the available information.

---

[32]Speed [1985] considers the "Sizewell B Probabilistic Safety study" a more recent study that avoids some of the very worst of the abuses of the earlier report. He finds some of the basic criticisms still apply. Speed [1985] also references some of the literature on this subject.
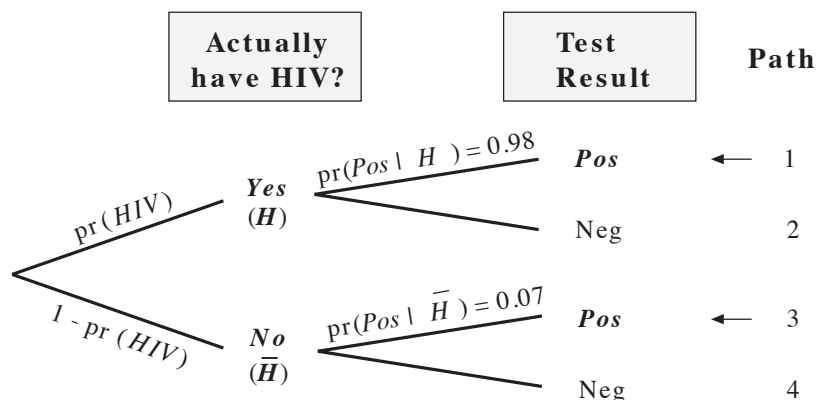[33]Recall from Case Study 4.6.2 that HIV is the virus that causes AIDS.

Figure 4.8.1 : Probability tree for HIV testing.

Since the required conditional probability is not immediately available to us, we expand it out using the conditional probability formula,

$$\text{pr}(HIV \mid Positive) = \frac{\text{pr}(HIV \cap Positive)}{\text{pr}(Positive)} \,.$$

Having done this, we see that we everything we need can be read off the tree diagram (using the familiar rules of Section 4.6.3). No new theory is needed! The numerator comes from Path 1 of the tree, whereas we get the denominator by adding all paths for which the event *Positive* occurs, namely 1 and 3. Thus

$$\text{pr}(HIV \mid Positive) = \frac{\text{pr}(HIV) \times 0.98}{\text{pr}(HIV) \times 0.98 \;+\; (1 - \text{pr}(HIV)) \times 0.07} \,.$$

In the scenario above $\text{pr}(HIV) = 0.01$. If we substitute this value into the equation, we obtain

$$\text{pr}(HIV \mid Positive) \approx 0.12 \,.$$

The chances that a person who has tested positive really has the disease are not large. At approximately one chance in 8, they are actually moderately small.[34]

We shall discuss the practical implications of calculations such as this in Case Study 4.8.1. In the meantime, we concentrate on the calculations themselves. The method used in Example 4.8.1 also enables us to solve more complicated examples involving larger partitions of the sample space.

Example 4.8.2 Let us revisit the contraceptive failure problem of Example 4.6.7 in Section 4.6.3 where we had information on the proportions of women using the different methods of contraception, i.e. information of the

---

[34]Many people find results like this so counter-intuitive that they doubt the arguments. If that is the case for you, try this less technical argument. Suppose we had 10,000 people. We would expect 100 to have HIV (1%) and of this 100, 98 (98%) to test positively. We would expect 9,900 people out of the 10,000 (99%) not to have HIV and of these 9,900, we would expect 693 (7%) to test positively. This gives us 791 positive tests of which only 98, or 12% belong to people with HIV.

form pr(*Method*), and also on the failure rates of each method, i.e. of the form pr(*Failure* | *Method*).

One of the types of method was barrier methods. Suppose we wanted to obtain the proportion of just those experiencing contraceptive failure who were using barrier methods, i.e. pr(*Barrier* | *Failure*). Since the order of conditioning desired is the reverse of the order of conditioning in the available information, we expand it out using the conditional probability formula,

$$\text{pr}(Barrier \mid Failure) = \frac{\text{pr}(Barrier \cap Failure)}{\text{pr}(Failure)}$$

All of the information on the right hand side can be obtained from the tree in Fig. 4.6.4. The numerator is obtained from Path 5 in which both *Barrier* and *Failure* occur giving pr(*Barrier* ∩ *Failure*) = 0.24 × 0.14. We can find pr(*Failure*) by adding the probabilities of all paths containing *Failure*. We did this in Example 4.6.7 and found that pr(*Failure*) = 0.0592. Thus

$$\text{pr}(Barrier \mid Failure) = \frac{\text{pr}(Barrier \cap Failure)}{\text{pr}(Failure)} = \frac{0.24 \times 0.14}{0.0592} = 0.568\,.$$

Almost 60% of those with unwanted pregnancies relied primarily on barrier methods of contraception.

In the above example, we have used the tree diagrams to find probabilities. However, another approach is possible. As an exercise, the reader might like to try and construct Table 4.8.1.

Table 4.8.1 : Table Constructed from the Data in Example 4.6.7

	Steril.	Oral	Barrier	IUD	Sperm.	Total
Failure (F)	0	.016	.0336	.0018	.0078	.0592
Success (S)	.38	.304	.2064	.0282	.0222	.9408
Total	.38	.320	.2400	.0300	.0300	1.0000

The entry in the first row and third column of Table 4.8.1, namely the probability pr(*Barrier* ∩ *Failure*), is given above. This number can be arrived at intuitively by noting that the 14% failure rate applies to 24% of the women giving 0.14 × 0.24 = 0.0336. The remaining entries follow in a similar fashion. Adding up the numbers in the first row gives pr(*F*) = 0.0592, the answer to Example 4.6.7. Furthermore, pr(*Barrier* | *Failure*) can be found by dividing the entry in the table (0.0336) by the first row sum (.0592) to get 0.568, the answer to Example 4.8.2. We have demonstrated that the construction of such tables provides an alternative way of finding various probabilities, including conditional probabilities. ∎

Bayes Theorem

(i) $\mathrm{pr}(B \mid A) = \dfrac{\mathrm{pr}(A \mid B)\mathrm{pr}(B)}{\mathrm{pr}(A)} = \dfrac{\mathrm{pr}(A \mid B)\,\mathrm{pr}(B)}{\mathrm{pr}(A \mid B)\,\mathrm{pr}(B) + \mathrm{pr}(A \mid \overline{B})\,\mathrm{pr}(\overline{B})}.$

(ii) If $C_1, C_2, \ldots, C_k$ is a partition of S

$$\mathrm{pr}(C_i \mid A) = \dfrac{\mathrm{pr}(A \mid C_i)\mathrm{pr}(C_i)}{\mathrm{pr}(A)} = \dfrac{\mathrm{pr}(A \mid C_i)\,\mathrm{pr}(C_i)}{\sum\limits_{j=1}^{k} \mathrm{pr}(A \mid C_j)\,\mathrm{pr}(C_j)}.$$

Note that item(i) below is a special case of item (ii) in which the partition used consists the two events B and $\overline{B}$.

[***Proof of (ii):*** That the numerator of the righthand-side gives $\mathrm{pr}(C_i \cap A)$ follows from the multiplication rule; note that $\mathrm{pr}(C_i \cap A)$ is identical to $\mathrm{pr}(A \cap C_i)$. The fact that the denominator sums to give $\mathrm{pr}(A)$ is the alternative expression of the Partition Theorem given in Section 4.6.3].

The steps we used in the examples to find $\mathrm{pr}(\textit{Event 1} \mid \textit{Event 2})$ when the order of conditioning required is "reversed" from that in the available information were:

1. Put the relevant available information on a tree diagram.

2. Expand $\mathrm{pr}(\textit{Event 1} \mid \textit{Event 2}) = \dfrac{\mathrm{pr}(\textit{Event 1} \cap \textit{Event 2})}{\mathrm{pr}(\textit{Event 2})}$.

3. Obtain the required probabilities from the tree.

For this method to work, it is critical that the first set of branches in the tree forms a partition of the sample space (in the case of probabilities) or population of interest (in the case of proportions). Equivalently, the probabilities on the first fan of branches must add to unity.

We saw in Example 4.8.2 above that such "reversed" probabilities can also be found by constructing a two-way table of probabilities.

*Case Study 4.8.1   AIDS, Lie detectors and Job Competency

The ELISA test for HIV infection was described in Case Study 4.6.2 and further discussed in Example 4.8.1. It correctly classifies the vast majority of infected people as having HIV. It also correctly classifies the vast majority of uninfected people as not having HIV. And yet Example 4.8.1 showed a scenario in which the majority of people testing positive for HIV were in fact uninfected.

This sort of "good-but-imperfect-test" situation is widespread. It applies to large numbers of medical screening procedures (diabetes, cervical cancer, breast cancer, ...).[35] It applies to polygraph lie detector tests (some people who are not lying show the physiological symptoms interpreted as a sign of lying, while some people who are lying do not). It applies to psychological and intellectual tests performed to judge the suitability of job applicants (some

---

[35] However, a screening test is designed to identify a group at increased risk of a condition.

people who are capable of doing the job well will fail the tests, while some who are not will pass the tests). It can also apply to the testing of urine or blood samples to detect drug use. In this study we shall discuss some important problems associated with using and interpreting the results of such tests. The vehicle for our discussion is the ELISA test for HIV, but essentially the same considerations apply to all such tests.

There has been some popular pressure for mass screening for HIV. However, as we know from Case Study 4.6.2 and Example 4.8.1, with any such screening, there will be large numbers of people without HIV who turn up a positive ELISA test. This leaves you with the huge practical problem of identifying the minority of these people who are actually infected. Let us investigate the extent of the problem.

From Example 4.8.1 (and Fig. 4.8.1)

$$\text{pr}(HIV \mid Positive)$$
$$= \frac{\text{pr}(HIV)\,\text{pr}(Positive \mid HIV)}{\text{pr}(HIV)\,\text{pr}(Positive \mid HIV)\ +\ \text{pr}\,(No\ HIV)\,\text{pr}(Positive \mid No\ HIV)}$$
$$= \frac{\text{pr}(HIV) \times 0.98}{\text{pr}(HIV) \times 0.98\ +\ (1 - \text{pr}\,(HIV)) \times 0.07}\,. \tag{1}$$

In this context, $\text{pr}(HIV)$ is the proportion of people in the population being screened who actually have *HIV*. Table 4.8.2 gives estimated $\text{pr}(HIV)$ values for several different countries (column 4) and the resulting value of $\text{pr}(HIV \mid Positive)$, the proportion of those testing positive who really have HIV (final column).

Table 4.8.2 : Proportions Infected with HIV

Country	No. AIDS[a] Cases	Population[b] (millions)	$\text{pr}(HIV)$[c]	$\text{pr}(HIV \mid Positive)$
United States	218,301	252.7	0.00864	0.109
Canada	6,116	26.7	0.00229	0.031
Australia	3,238	16.8	0.00193	0.026
New Zealand	323	3.4	0.00095	0.013
United Kingdom	5,451	57.3	0.00095	0.013
Ireland	142	3.6	0.00039	0.005

[a]Source: AIDS - New Zealand, November 1992.
[b]1991 estimates – except for Ireland for which May 1990 figures are given.
[c]Proportion of population infected by HIV. These are very rough. We have assumed that the proportion of HIV infected people is 10 times larger than the proportion of AIDS cases. This is the approximate relationship between the number of US cases and the US Centers for Disease Control's estimate of the number of HIV infected Americans in 1990.

The most extreme case in Table 4.8.2 is Ireland. If the Irish Government had decided to screen the total population of 3.6 million people for HIV in 1990,

on the figures above, roughly $250,000$ (7%) would have tested positively and of these only about 1250 (0.5% of the positives) would have HIV. How do we tell these 1250 people apart from the rest of the $250,000$? In the case of HIV there is another more expensive and more specific test called the Western Blot test which could be used.[36] So, any screening program would have to include funding for both ELISA tests for everyone and Western Blot tests for a quarter of a million people.[37]

Although the value of $\text{pr}(HIV \mid Positive)$ in Table 4.8.2 varies with the proportion of people with HIV in the population to some extent, all of the entries in the table are small. We don't want to leave you with the reverse misapprehension that $\text{pr}(HIV \mid Positive)$ is always small. Among intravenous drug users in New York in 1988, it was estimated that 86% had HIV (*NZ Herald*, 17 November 1988). Using $\text{pr}(HIV) = 0.86$ in equation (1) above, we now find that $\text{pr}(Positive) = 0.853$ and $\text{pr}(HIV \mid Positive) = 0.988$. If all New York drug addicts had been screened almost every person testing positively (98.8%) would have had HIV.

The type of problem we see in Table 4.8.2, where the majority of those that would test positive would be false positives, is very common in screening for relatively rare medical conditions. Similar behavior could be expected in testing for drug use among a population in which drug use is rare, or using lie detector tests on a group of people in which the vast majority had told the truth. An alternative strategy, as indicated by the results for New York drug addicts, is to try to identify high risk subpopulations and only screen those. With medical screening, particularly in an area as sensitive as AIDS, this can be political dynamite.

So far, we have been using $\text{pr}(HIV \mid Positive)$ to think about the proportion of those testing positive in a screened population actually have HIV. But what does $\text{pr}(HIV \mid Positive)$ mean for an individual?

Let's get personal and imagine that you, the reader, have just tested positive. Clearly, this would be a major trauma for you. *TIME* (2 March 1988) quoted a health professional as saying, "The test tends to rip people's lives apart". The *Economist* (4 July 1992) told a story of a young American having recently committed suicide on learning that he had tested positive for HIV. "...he believed his chances of carrying the virus was 96%. It was 10%".[38] So, what is $\text{pr}(HIV \mid Positive)$ for me, i.e. what is *the probability that **I** have HIV given that **I** have just tested positive* on an ELISA test?

---

[36] In medical terms (see footnote, Case Study 4.6.2) the Western blot is more specific but not as sensitive as Elisa.

[37] Such testing is not cheap! The State of Illinois introduced screening as a condition for a marriage license in 1988. In the first 11 months $150,000$ people were screened at a cost of US$5.5 million (23 were infected). Many other states now do similar screening.

[38] It is surprising that someone was given the results of a positive result on a single test. In NZ, people are not told that they have tested positive unless they have also tested positive on a second ELISA test and on a Western blot test.

We have to think in terms of being a random representative of some population. We saw above that the value of pr(*HIV* | *Positive*) depends critically on the value of pr(*HIV*) for the population the individual is sampled from. None of us can usefully be thought of as a randomly selected individual from our own country as far as HIV is concerned because we know that HIV is much more prevalent in some sections of the population than others. To obtain a value of pr(*HIV* | *Positive*) for oneself, a value for pr(*HIV*) is required that gives the proportion of people who have HIV *among people as much as possible like oneself* with respect to the known risk factors for AIDS. If you are a New York drug addict who shares needles, a positive ELISA test is fairly conclusive. If you have always lived in a monogamous sexual relationship, believe your partner to have done the same, don't share needles, and didn't have a blood transfusion prior to the testing of the blood supply, a positive ELISA test is almost certainly a false positive.

Exercises on Section 4.8

1. Suppose, as above, the ELISA test gave a positive result on a "healthy" individual with probability 0.07 and a positive result on a HIV carrier with probability 0.98. Suppose we did the test twice and concluded the person had HIV only if both tests were positive.

   (a) Assuming test results are independent, what is the probability of saying: **(i)** a healthy person has HIV?, **(ii)** a person with HIV does not have the disease?

   (b) Repeated testing looks like a way of dramatically cutting down the false positive rate. However in Case Study 4.6.2 it was stated that the test reacts to other antibodies besides HIV. What are the consequences of this to the independence assumption?

2. If you only wanted to use the ELISA test to protect the blood banks, what sort of cutoff MAR value would you use from Table 4.6.3? Why? [Think of the consequences of the two types of wrong decision.]

3. 43% of the North Korean workforce works in agriculture versus 21% of the South Korean laborforce (*TIME*, 2 July 1990). 34.8% of the Korean workforce lives in the North and the rest in the South. By constructing a two-way table of probabilities, find the proportion of Korean agricultural workers that live in the North.

Many of the Review Exercises for this chapter involve reversing conditional probabilities.

4.9 Summary

From Chapter 4, we hope you absorb:

(a) Some basic ideas about probabilities, such as: how probabilities arise; the idea of a simple probability model; the important (but very different) notions of mutually exclusive events and (statistically) independent events; the idea of

assessments of probabilities depending upon the information available and a formalization of this through conditional probability.

(b) Some facility with manipulating probabilities. There are some standard types of probability manipulation that will be used repeatedly, particularly in the following chapter. The adding of probabilities of mutually exclusive events (to get the probability that one of them occurs) and multiplying the probabilities of independent events (to get the probability that all of them occur), in particular, fall into this class. Also some problems can best be solved by constructing two-way tables of counts or proportions.

You will find that many of the very simple examples that you have used and thought about in this chapter (e.g. tossing coins, or sampling colored balls) will become models (analogies) which will enable you to solve some practically important problems in the following chapters. Conditional probabilities play no further part in the book.[39] They are included here because of their importance in thinking about a number of very important practical problems (as illustrated in the Case Studies) and to combat the widespread misuse of probability arguments, particularly those based upon assumptions of independence.

This summary has been divided into two sections, the first dealing with the main concepts or ideas about probability, and the second dealing with formulae for calculating and manipulating probabilities.

4.9.1 Summary of concepts

1. The ***probabilities*** people quote ***come from 3 main sources***:
    (i) ***models*** (idealizations like the notion of equally likely outcomes which suggest probabilities by symmetry).
    (ii) ***data*** (e.g. relative frequencies with which the event has occurred in the past).
    (iii) ***subjective feelings*** representing a degree of belief.

2. A simple ***probability model*** consists of a *sample space* and a *probability distribution* (definitions to follow).

3. A ***sample space***, *S*, for a random experiment is the set of all possible outcomes of the experiment.

4. A list of numbers $p_1, p_2, \ldots$ is a ***probability distribution*** for a discrete sample space $S = \{s_1, s_2, s_3, \ldots\}$ provided (i) all of the p_i's lie between 0 and 1, and (ii), they add to 1.

    According to the probability model, p_i is the probability that outcome s_i occurs. We write $p_i = \mathrm{pr}(s_i)$.

5. An ***event*** is a collection of outcomes.

---

[39]They do provide an alternate means of answering many of the problems in the next chapter, however.

An event ***occurs*** if any outcome making up that event occurs.

6. The ***probability of event A*** can be obtained by adding up the probabilities of all the outcomes in A.

7. If all outcomes are equally likely,

$$pr(A) = \frac{\text{Number of outcomes in } A}{\text{Total number of outcomes}}$$

8. The ***complement*** of an event A, denoted $\overline{A}$, occurs if A does not occur.

9. It is useful to represent events diagrammatically using so-called ***Venn diagrams***.

10. ***Unions*** of events: $A \cup B$ contains all outcomes in A or B (including those in both). It occurs if at least one of A or B occurs.

11. ***Intersections*** of events: $A \cap B$ contains all outcomes which are in **both** A and B. It occurs only if both A and B occur.

12. ***Mutually exclusive*** events cannot occur at the same time.

13. A ***partition*** is a way of dividing up a sample space into separate pieces. Events $C_1, C_2, \ldots, C_k$ form a partition of the sample space if they are mutually exclusive and collectively account for all possible outcomes.

14. The (***conditional***) ***probability*** of A occurring ***given*** that B occurs is given by

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

15. Events A and B are (***statistically***) ***independent*** if knowing whether B has occurred gives no new information about the chances of A occurring, i.e. if $pr(A \mid B) = pr(A)$.

16. If events are physically independent, then, under any sensible probability model, they are also statistically independent.

17. Assuming that events are independent when in reality they are not can often lead to answers that are grossly too big or grossly too small.

4.9.2 Summary of useful formulae

A. For discrete sample spaces, $\mathbf{pr(A)}$ can be obtained by adding the probabilities of all outcomes in A.

B. **For equally likely outcomes** in a finite sample space

$$pr(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}.$$

C. General Probability Rules.[40]

1. $\operatorname{pr}(S) = 1$, $\operatorname{pr}(\varnothing) = 0$.

2. $\operatorname{pr}(\overline{A}) = 1 - \operatorname{pr}(A)$.

3. $\operatorname{pr}(A \cup B) = \operatorname{pr}(A) + \operatorname{pr}(B) - \operatorname{pr}(A \cap B)$
   $\left(\text{if } A \text{ and } B \text{ are mutually exclusive } \operatorname{pr}(A \cap B) = 0\right)$.

4. (a) $\operatorname{pr}(A) = \operatorname{pr}(A \cap B) + \operatorname{pr}(A \cap \overline{B})$
   $= \operatorname{pr}(B) \operatorname{pr}(A \mid B) + \operatorname{pr}(\overline{B}) \operatorname{pr}(A \mid \overline{B})$.

   (b) If $C_1, \ldots, C_k$ form a partition

$$\operatorname{pr}(A) = \sum_{i=1}^{k} \operatorname{pr}(A \cap C_i) = \sum_{i=1}^{k} \operatorname{pr}(C_i) \operatorname{pr}(A \mid C_i).$$

D. Conditional Probability.

1. Definition

$$\operatorname{pr}(A \mid B) = \frac{\operatorname{pr}(A \cap B)}{\operatorname{pr}(B)}.$$

2. Multiplication formula

$$\operatorname{pr}(A \cap B) = \operatorname{pr}(A) \operatorname{pr}(B \mid A) = \operatorname{pr}(B) \operatorname{pr}(A \mid B).$$

$$\operatorname{pr}(A_1 \cap A_2 \cap \ldots \cap A_n) = \operatorname{pr}(A_1) \operatorname{pr}(A_2 \mid A_1) \ldots \operatorname{pr}(A_n \mid A_1 \cap \ldots \cap A_{n-1}).$$

E. Independence

If $A_1, \ldots, A_n$ are mutually independent, it follows that

$$\operatorname{pr}(A_1 \cap A_2 \cap \ldots \cap A_n) = \operatorname{pr}(A_1) \operatorname{pr}(A_2) \ldots \operatorname{pr}(A_n).$$

Review Exercises 3

1. Give a suitable sample space S for each of the following random experiments. (Note that there may be more than one answer for S.)

(a) A light bulb is chosen at random from a batch of bulbs. It either works or it doesn't.

(b) A student is selected at random from your class and their number of siblings (brothers and sisters) is recorded.

(c) A person is interviewed on the street and the number of their parents that are alive is noted as part of a questionnaire.

(d) You have a thermometer hanging at home and you read the temperature at a given time every day.

---

[40] As demonstrated in the chapter, a diagram or table will often give you the probabilities that you need without you having to remember or look up these formulae.

(**e**) One lunchtime you go and count the number of people in the queue at the student cafeteria.

(**f**) A member of your class of 10 students is chosen at random and their height is measured.

(**g**) A randomly selected student is interviewed and they are asked what form of transport they used to get to the university that day.

2. Let $S = \{s_1, s_2, s_3\}$.

(**a**) Determine whether or not S is a sample space in the following cases:

(**i**) $s_1 =$ "My alarm goes off and I get to my lecture on time".
$s_2 =$ "My alarm doesn't go off and I get to my lecture".
$s_3 =$ "I miss my lecture or arrive late".

(**ii**) $s_1 =$ "I catch my bus and get to my lecture on time".
$s_2 =$ "I miss my bus and miss my lecture."
$s_3 =$ "I get to my lecture on time".

(**b**) Describe a random experiment underlying this exercise.

3. According to the 1991 NZ census, a randomly selected dwelling would have 1, 2, ... occupants with probabilities given by the following table:

No. of occupants	1	2	3	4	5	6 or more
Probability	.21	.32	.17	.16	—	.05

Source: Statistics NZ, 1994.

Obtain the missing probability corresponding to 5 occupants.

4. An experiment has two outcomes A and $\overline{A}$. If A is three times as likely to occur as $\overline{A}$, what is $\mathrm{pr}(A)$?

5. Suppose we roll a normal six-sided die and observe the score. The sample space for this random experiment is $S = \{1, 2, 3, 4, 5, 6\}$. Define the events $A = \{1, 4\}$, $B = \{6\}$, $C = \{1, 2, 3\}$ and $D = \{2, 3, 5\}$.

(**a**) What is the event E that the score is an odd number.

(**b**) What is the complement of D?

(**c**) Which of the events A to E form a partition of S?

(**d**) Which of the events B to E are mutually exclusive to A?

(**e**) Find the following events: (i) $A \cup C$ (ii) $C \cap D$ (iii) $C \cap \overline{D}$.

6. Craps is a dice game played in casinos, among other places. If you see people rolling dice along a floor in a movie, the game being played is probably craps. The player rolling the dice is called the "shooter". The shooter plays by rolling a pair of dice. The score for each roll is the sum of uppermost faces of the two dice. Initially the shooter rolls the pair of dice. If the shooter rolls a "natural", i.e. scores a 7 or an 11 the shooter wins outright. If the shooter throws "craps", i.e. a 2, 3 or 12 he or she

loses outright. Suppose the shooter throws any other score on the first roll, for example a 5. This score is called the shooter's "point". He or she continues to roll again and again until either this score is repeated (in which case the shooter wins or makes "point") or a seven is thrown (in which case the shooter loses).

We shall lead you through the calculation of the probability that the shooter wins in a way that revises many of the ideas in the chapter.

(**a**) Write down the usual sample space for rolling a pair of dice once.

(**b**) In terms of the outcomes in this sample space write down the following events: **(i)** $A_7 =$ "the total score is 7", **(ii)** $N =$ "a natural is thrown", and **(iii)** $C =$ "the rolling results in craps".

(**c**) Write down the probabilities of the events in (b) and also the probability of $F =$ "continue rolling".

(**d**) State why $\{A_7,\ N,\ F,\ C\}$ does not form a sample space for the two dice experiment and modify it so that it does.

Now let us start doing the calculations. $S = \{2,\ 3, \dots, 12\}$ forms a sample space for the experiment formed from just using the total score on the initial roll of the two dice.

(**e**) Write down the probability distribution corresponding to this sample space.

Let A_i be the event that the total score from that first roll of the pair of dice is i, for $i = 2,\ 3, \dots, 12$ (e.g. A_8 is the event that the score from the first roll, or point, is 8).

(**f**) We now want to calculate all the probabilities of the form pr("shooter wins" given that the initial score was i), i.e. pr("shooter wins"$\,|\,A_i$).

  (**i**) For $i = 2,\ 3,\ 7,\ 11$ and 12 we can write down the probabilities immediately from the description of the game without doing any calculations. Write down these 5 probabilities.

  (**ii**) Suppose the shooter's point was 4. There are 3 outcomes giving rise to a total score of four and 6 outcomes giving rise to a score of seven. It can be shown that the chances of another four turning up before a seven are in the same ratio, i.e. $3:6$. Thus the probability of a 4 turning up before a 7 is $3/(3+6) = 3/9$. Thus

  $$\text{pr("shooter wins" } |\ A_4)\ = 3/9.$$

  Using the same idea write down all the remaining conditional probabilities $P(\text{"shooter wins" } |\ A_i)$.

  [Note: These probabilities are important because the gamblers are free to bet at any stage of the game on whether the shooter will make point.]

(**g**) By noting that the events A_i define a partition of possible games, calculate the probability that the shooter wins?

(**h**) Suppose the shooter is "hot", e.g. has won eight straight games. What is the probability that the shooter wins the next game? Why?

[Epilogue: According to Larsen and Marx [1986, page 72], in 1980 a man walked into the Horseshoe Club in downtown Las Vegas carrying two suitcases. One was empty and the other contained US$770,000 in $100 bills. He went over to a craps table and bet the lot against the shooter. The woman who was shooting rolled a 6 then a 9 and then a 7 (and lost!). The man then left the casino after visiting the cashier with both suitcases full of money. As of 1986, this was the largest single bet ever recorded in Las Vegas. A note of caution – he had an almost 50 : 50 chance of walking out with two empty suitcases.]

7. One of the biggest problems with conducting a mail survey is the poor response rate. In an effort to reduce nonresponse, several different techniques for formatting questionnaires have been proposed. An experiment was conducted to study the effect of the questionnaire layout and page size on response in a mail survey. A group of students at a Dutch university were questioned about their attitudes towards suicide. Four different types of questionnaire formats were used. The results of the survey are shown in the Table 1.

Table 1 : Questionnaire Formats[a]

Format	Responses	Nonresponses	Total
Typewritten (small page)	86	57	143
Typewritten (large page)	191	97	288
Typeset (small page)	72	69	141
Typeset (large page)	192	92	284
Total	541	315	856

[a]Source: Jansen [1985].

(a) What proportion of the sample responded to the questionnaire?

(b) What proportion of the sample received the typeset (small page) version?

(c) What proportion of those who received a typeset (large page) version actually responded to the questionnaire?

(d) What proportion of the sample received a typeset (large page) questionnaire and responded?

(e) What proportion of those who responded to the questionnaire actually received a typewritten (large page) questionnaire?

(f) By looking at the response rates for each of the four formats, what do you conclude from the study?

8. French gamblers in the seventeenth century used to bet that at least one "one" would turn up on four rolls of a die. A "one" is often called an ace. In a similar game they also bet on whether at least one double ace would occur on 24 rolls of a pair of dice. A French nobleman, the Chevalier de Méré, reasoned that both events were equally likely arguing essentially as follows:

Game 1: There is 1 chance in 6 of getting an ace on one roll so in four rolls, the chances are four times that i.e. $4/6 = 2/3$.

Game 2: There is a chance of $1/36$ of getting a pair of aces from rolling a single pair so for 24 pairs, the chances are 24 times that or $2/3$ again.

However, from experience, de Méré doubted his arguments. It seemed to him that in practice the first event occurred more frequently. This situation came to be called the Paradox of Chevalier de Méré.

De Méré passed his problem on to Blaise Pascal (1623–1662). It was solved in a correspondence between two of the greatest mathematical minds of the day, Pascal himself and Pierre de Fermat (1601–1665). With your understanding of probability theory the problem should not be very difficult. However Pascal and Fermat were starting from scratch with very few of the probabilistic ideas that make thinking about such problems fairly easy for us. In fact their correspondence was the starting point for the development of a mathematical theory of probability.

(a) What is wrong with Chevalier de Méré's arguments? (Hint: for Game 1 apply the argument when you use only 2 rolls.)

(b) What is the true probability in either case?

[Note: Although the probabilities are different, they are very similar. Later we shall do some calculations to get some idea of the immense number of games one would have to watch to detect fairly reliably that these probabilities are in fact different.]

9. The University of Auckland has Faculties as follows (with the percentage of the student body in each following in parenthesis): Arts (30%), Commerce (19%), Science (18%), Engineering (7%), Law(7%), Education (6%), Medicine(4%) and Other (9%), where "Other" encompasses the remaining Faculties which are smaller than those listed. The percentages of female students within these Faculties are: Arts (65%), Commerce (41%), Science (39%), Engineering (15%), Law(52%), Education (82%), Medicine(49%) and Other (47%).

(a) Construct a two-way table showing the percentages of males and females in the various faculties.

(b) What percentage of Auckland's students are female?

(c) What percentage of Auckland's female students are in **(i)** Arts? **(ii)** Law? **(iii)** Engineering? **(iv)** Education?

10. Arab citizens make up 14% of the population of Israel. Also 11% of Israel's Jews and 52% of its Arab citizens live below the poverty line (*TIME*, 13 April 1992). Assume that Jews and Arabs account for the whole population of Israel.

(a) What proportion of Israel's population lives below the poverty line?

(b) What proportion of Israel's poor (i.e. those below the poverty line) are Arab?

(c) Set up an appropriate two-way table and use it to answer (a) and (b) again.

11. In 1980, a US Senate Committee was investigating the feasibility of a national screening program to detect child abuse. A team of consultants came up with the following estimates: 1 child in 100 is abused; a physician can detect an abused child about 90% of the time; and a national screening program using physicians would incorrectly label about 3% of the nonabused children as abused.

(a) Using the above information, what is the probability that a child is actually abused given the screening program diagnoses the child as such?

(b) Do you think it is appropriate to apply the above information to both boys and girls?

(c) Would the above information be relevant today?

12. There is a 40% to 60% chance that a pregnant woman with the HIV virus will pass it on to her child. Approximately 1% of all black teenage girls who bore children in New York City during 1988 were infected with HIV (*TIME*, 2 July 1990). Taking the lower figure (40%), what proportion of the babies born to black teenage girls in 1988 were infected? Express your answer as a rate of infected babies per thousand births.

13. Most of the figures to follow come from a *TIME* cover story (6 July 1992) about effective new drugs for treating schizophrenia. All are US figures and all are approximate estimates.

(a) One in four schizophrenics attempts suicide. Of those who attempt it, one in 10 succeeds. What proportion of schizophrenics actually commit suicide?

Approximately 1% of of the population is schizophrenic, 0.8% of people are homeless, and one third of the homeless are schizophrenic.

(b) What proportion of schizophrenics are homeless? (Difficult. You cannot do this on a single tree.)

A child has a 10% chance of becoming schizophrenic if one parent is schizophrenic and a 40% chance if both are. The chance of a child of a schizophrenic developing the condition is reduced, but only slightly, when raised by adoptive parents without the condition.

(c) Does this information suggest anything to you about hereditary and environmental effects for schizophrenia?

(d) It would be interesting to know what percentage of schizophrenics had schizophrenic parents. What additional information would we need to resolve this?

(e) Assuming that people marry independently of their susceptibility to schizophrenia and each couple has one child, answer the question raised in (d).

(f) Actually our "1% of Americans are schizophrenic" is not quite correct. The article really says that 1% of Americans *develop* schizophrenia. What can you therefore say about the percentage who *have* schizophrenia?

14. The final grade of students in a first year course in Statistics at Auckland was made up of 30% from assignments and tests while the course is in progress (to form a "coursework mark") and 70% from an examination at the end of the course. Table 2 showing the relationship between course-work mark and final grade is taken from the results of 1270 students who completed the course in 1992. The entries in the interior of the table are the percentages of those students within the given range of coursework marks who got each of the final grades.

(**a**) Can you think of a reason for presenting the data in this way (or a use for it as presented)?

Table 2 : Coursework Marks and Final Grades

Percentage of Class	Coursework Mark	Final Grade				Total
		Fail Grade	C	B	A	
2.7	≤ 5	91.2%	8.8%	0	0	100%
4.6	$5^+ - 10$	79.7	20.3	0	0	100
18.1	$10^+ - 15$	53.9	35.7	10.0	0.4	100
33.9	$15^+ - 20$	13.3	45.1	34.9	6.7	100
30.6	$20^+ - 25$	0.3	12.6	60.3	26.8	100
10.1	$25^+ - 30$	0	0	14.7	85.3	100
100						

Suppose that we randomly select a student who completed the course.

(**b**) Describe, in words, a sample space for this experiment.
(**c**) What is the probability that the student got between 25 and 30 for coursework?
(**d**) What is the probability that the student got between 25 and 30 for coursework and an A-grade?

Suppose that the figures in Table 2 lead to reasonable probabilities for a randomly selected student currently taking the course.

(**e**) What is the probability the student will get an A-grade given that he or she got between 25 and 30 for coursework?
(**f**) What is the probability that a student with between 10 and 15 for coursework will get a B-grade or better?

The remainder of this exercise talks about proportions of the 1992 class.

(**g**) What proportion of the class got between 15 and 20 on coursework?
(**h**) What proportion of the class got between 15 and 20 on coursework and failed?
(**i**) What proportion of the class with between 15 and 20 on coursework failed?
(**j**) What proportion of the entire class failed? What proportion passed?
(**k**) What proportion of the class got A-grades?
(**l**) What proportion of those with A-grades got over 25 on coursework?

15. In North America, as in Australasia, cancer is the second leading cause of death after heart diseases. Accidents account for only about a fifth as many deaths as cancer. Table 3 gives the incidence rates (as new cases per hundred thousand of population) and the mortality rates (as deaths per hundred thousand) for 7 leading cancer sites.

Table 3 : Cancer Incidence and Mortality Rates[a]

Cancer Sites	New Cases			Deaths		
	Male	Female	Total	Male	Female	Total
Oral	20.6	9.7	30.3	5.2	2.8	8.0
Colo-Rectum	79.0	77.0	156.0	28.9	29.4	58.3
Lung	102.0	66.0	168.0	93.0	53.0	146.0
Skin[b]	17.0	15.0	32.0	4.1	2.6	6.7
Breast	1.0	180.0	181.0	0.3	46.0	46.3
Uterus	0	45.5	45.5	0	10.0	10.0
Prostate	99.0	0	99.0	28.0	0	28.0
Other	246.4	171.8	418.2	115.5	101.2	216.7
Total	565.0[c]	565.0[c]	1130.0	275.0	245.0	520.0

[a]Source: Constructed from data in The World Almanac and Book of Facts [1993, page 223] except for prostate figures which are 1988 figures from American Cancer Soc.
[b]Excludes non-melanoma skin cancer.
[c]These are the same simply because of the rounding of the data.

(a) What cancer has most new cases in a year?

(b) What cancer kills: **(i)** the most people? **(ii)** the most males? **(iii)** the most females?

(c) What proportion of breast cancer deaths are males?

(d) What proportion of male cancer deaths are due to: **(i)** lung cancer? **(ii)** colon/rectum cancer?

(e) What listed cancers affect: **(i)** males and not females? **(ii)** females and not males?

The numbers of cancers contracted in a year and the numbers of deaths in a year from that cancer give us a rough estimate of the chances that a particular cancer will eventually kill someone who contracts it.

(f) Which listed cancer is most likely to end up being fatal: **(i)** regardless of gender? **(ii)** for men? **(iii)** for women? In each case, give the probability of eventual death.

(g) What assumptions underly the type of calculations done in (f)? When are these assumptions sensible?

(h) Doctors are more likely to talk about things like the 5-year survival-rate than about the chances of eventual death from a disease. What are some problems with "the chance that you will die from it" as a measure of the seriousness of a disease?

16. The 15,679 known HIV cases in Australia up until December 1991, were classified according to the origin of the infection as given in Table 4.

Table 4 : HIV Cases in Australia to December 1991[a]

Category	Number
Male homosexual/bisexual contact only	7,597
Male homosexual/bisexual contact and intraveneous drug use	248
Intraveneous drug use only	499
Heterosexual contact	407
Haemophilia/coagulation disorder	209
Received from blood products or tissue (other than haemophiliac etc)	139
Other/undetermined	6,488
Children	92
Total	15,679

[a]Source: Australian HIV Surveillance Report as reported by *TIME*, 20 April 1992.

(a) The categories in the list were intended to form a partition of HIV sufferers. How do we know this?

There were ambiguities in the original wording of the list: the word "only" was added by us to try and reduce the potential confusion. We shall define new events (or categories):

> Hom = "male homosexual/bisexual contact",
> $IVUse$ = "intravenous drug use",
> $Hetero$ = "heterosexual contact",
> $Haem$= "haemophilia/coagulation disorder", and
> $Blood$ = "received infected blood products".

(b) By inspecting the table, which of these events are mutually exclusive and which overlap?

(c) There is no apparent overlap between *Hetero* and *Hom*, or *Hetero* and *IVUse*. What does this mean?

(d) Represent each of the first 6 categories listed in the table above in terms of the new events.

(e) Represent the information in the table using a Venn diagram based upon the newly defined classes and include on your diagram the numbers of HIV cases given in the table.

(f) What proportion of all HIV sufferers is known to fall into each of the newly defined categories *Hom* to *Blood*, i.e. what is pr(*Hom*), pr(*IVUse*), etc?

(g) Is heterosexual contact a more common source of HIV infection than receipt of blood products?

(h) What proportion of HIV sufferers is known to have contracted the virus by homosexual/bisexual contact or intravenous drug use?

(i) Among all properly classified HIV sufferers (i.e. those not in the catch all "Other/undetermined" category) what proportions are in each of the categories *Hom* to *Blood*?

(j) What proportion of those infected by receiving blood products, had haemophilia or a coagulation disorder?

17. Reporting on an address by Professor William Doe, President of the Gastroenterological Society of Australia, *The Weekend Australian* (30 June 1990) stated that tests used widely throughout Australia to detect bowel cancer could lead to death because of widely inaccurate results. The tests, which used chemicals to detect the presence of blood in faeces, were available over the counter at pharmacies and widely used for mass screening programs. According to Professor Doe, they delivered a negative result on 30% to 50% of cancers (we shall assume 40%) and "these people are tragically assured that they don't have cancer". A positive result is delivered to 40% of people who do not have bowel cancer or its early warning signs, some of these resulting from animal blood in the bowel from digested meat. There are about 8,000 new cases of bowel cancer per year in Australia out of a population of approximately 16 million. We shall therefore estimate the proportion of undiagnosed cases of bowel cancer and its precursors in Australia at the time as approximately 0.0005. If a mass screening of all Australians was undertaken using this test:

(a) What proportion of bowel cancer cases would be overlooked?

(b) What proportion of Australians would give a positive test?

(c) What proportion of people testing positively would actually have the disease?

(d) How could you reduce the false positive rate?

[Note: The good news is that more reliable tests are now available.]

18. Predicting whether a paroled prisoner will go on to commit violent offenses is an extremely difficult task that no one seems to be able to do very well (Vasil [1987]). In a 1968 study, carried out by the California Department of Corrections, parolees were classified into two categories on the basis of offender histories and psychiatric reports. One fifth were classified as being "potentially aggressive" while the rest were classified as "less aggressive". In the potentially aggressive group, 3.1 per thousand were reconvicted for violent offenses committed within one year after release from prison. For the less aggressive group the rate was 2.8 per thousand.

(a) Of those reconvicted for violent offenses, what proportion had been classified "potentially aggressive"?

(b) What is notable about the above figures?

(c) The study does not strictly address the question of whether paroled prisoners *commit* violent offenses? What is it really measuring and what other factors are involved in that?

19. In 1993, a team of scientists from John Hopkins University and the University of Helsinki reported in *Science* (1993, vol. 260 p.751) their discovery of a genetic marker for so-called "familial cancer of the colon". This accounts for about one in seven cases of cancer of the colon (hereafter CaCo) which is the second leading cause of cancer deaths in the world. The scientists estimated that one person in 200 carries the defective gene, that 95% of people with the gene will develop cancer and, of those who get cancer, 60%

will get cancer of the colon.

(**a**) From these figures, what percentage of people will develop CaCo from this mechanism?

Professor Vogelstein of Johns Hopkins predicted a diagnostic test based upon the genetic discovery. (The existing screening test used world wide misses more than 70% of tumors.) Between five and 10 million Americans are presently considered to be at increased risk of CaCo because of a strong family history of the disease. Vogelstein believes that 75% of these people will find that they do not have the implicated genetic marker and that these people bear only the average risk of CaCo which is one chance in 20.

(**b**) What proportion of those with a "strong family history" will get CaCo? What proportion of those who will get CaCo carry the defective gene?

Detection of those carrying the marker is useful because their colons can then be scanned annually using a fiber-optic scope, a procedure known as colonoscopy which costs about $1,000$. (The five year survival rate for CaCo is 90% when detected early.)

(**c**) Taking the lower figure of 5 million Americans at increased risk, how much money could be spent annually on colonoscopies if Vogelstein's predictions are borne out?

20. Two tennis players play a 3 set match (i.e. they keep playing until one player wins two sets). Suppose that the two players are evenly matched so that each player has a 50% chance of winning a set and that the outcomes of different sets are independent.

(**a**) Write down a sample space for the experiment.

Let A denote the event "Player 1 wins the match" and B denote the event "Match finishes in 2 sets".

(**b**) Find **(i)** $\mathrm{pr}(A)$, **(ii)** $\mathrm{pr}(B)$, **(iii)** $\mathrm{pr}(A \cap B)$, **(iv)** $\mathrm{pr}(A \cup B)$, **(v)** $\mathrm{pr}(B \mid A)$.

(**c**) Are A and B independent? (Explain your answer.)

(**d**) Are A and B mutually exclusive?

21. Each day the price of a certain stock either moves up one cent or moves down one cent. It moves up with probability $1/3$ and down with probability $2/3$ independently of previous movements. We are interested in what the price will be in 3 days time. Let A be the event that the price has increased at the end of 3 days and let B be the event that the price drops on the first day. Write down a suitable sample space for this "experiment" and calculate:

(a) $\mathrm{pr}(A)$; **(b)** $\mathrm{pr}(B)$; **(c)** $\mathrm{pr}(A \cap B)$; **(d)** $\mathrm{pr}(A \mid B)$. **(e)** Are A and B **(i)** independent? **(ii)** mutually exclusive? Explain your answer.

22. Sanderson Smith [Smith, 1990] told the following story. He received an unsolicited phone call from a marketing research firm asking him to keep records and receipts for his purchases of certain items over a period of

three months. Some incentives were offered. The firm was to send him some journal booklets to record his purchases in; one journal for each of the three months. If he returned the journal for a given month along with receipts to verify his stated purchases he would go into that month's draw. Each draw would give him 1 chance in 10 of winning a prize of $25. Furthermore, if he returned all 3 journals and receipts on time, he would have 1 chance in 100 of winning a three day trip to Las Vegas, and 1 chance in 350 of winning a two-week holiday for two in Hawaii. Sanderson Smith did some calculations to figure out whether it was worth participating. Regarding the trips as the major prizes and the $25 prizes as minor, and assuming he fulfilled all the conditions, what is the probability of winning:

(a) **(i)** None of the minor prizes? **(ii)** All 3 minor prizes **(iii)** Exactly one minor prize?   **(iv)** Exactly 2 minor prizes?

(b) **(i)** Neither major prize? **(ii)** Las Vegas but not Hawaii? **(iii)** Both major prizes?

(c) **(i)** No prizes at all?   **(ii)** One or more minor prizes, but no major prize?

(d) If recording the data in the journals took 10 hours, would you be prepared to do it? [Sanderson Smith "declined to participate".]

23. Writing in *Rolling Stone* in February 1991, humorist P.J. O'Rourke ruminated on his experiences covering the Gulf War from Saudi Arabia. He pondered his chances of receiving a direct hit by a Scud missile. He figured that a Scud, which carried about 113 kg of explosive would have a blast area of 91 meters in diameter at most, and that missiles were being lobbed into an area of eastern Saudi Arabia that was roughly 80 km long and 48 km wide. Given that Scuds were inaccurately aimed, you can assume that missiles fall randomly on this area.

(a) What was his probability of being in the blast area of a missile?

(b) If 20 missiles fell onto the area, what is the probability of escaping all 20?

24. Suppose that an insurance company has classified its drivers into 3 classes using various criteria. Of drivers insured with them, 20% fall in the "low risk" category, 70% into "medium risk" category and 10% are "high risk". From the historical records, 1% of low risk drivers make a claim in any one year. Corresponding figures for medium risk and high risk drivers are 4% and 10% respectively. Assume that there have been no changes in the company's business that would prevent the historical record from being a good guide to the immediate future.

(a) Construct a two-way table of probabilities.

(b) What is the probability that a randomly selected driver in the medium risk category will have a claim?

(c) If we select a driver at random, what is the probability that the driver has a claim and is in the medium risk category?

(d) What proportion of drivers can be expected to have a claim (regardless of risk category)?

(e) What proportion of claims can we expect to be made by drivers classified as high risk?

*(f) If a driver goes 3 years without a claim, what is the probability he or she had been classified into the low risk group? (Treat different years as independent.)

25. In Chapter 1 of Brook et al. [1986], Barry Singer gives an entertaining and informative discussion of many of the ideas in this chapter and about finding a mate as well. This exercise revolves around your chances of finding an ideal mate.

(a) Make a list of the characteristics you consider absolutely essential in a partner (e.g. sufficiently intelligent, similar sense of humor, ...).

(b) Estimate or guess, for each item on the list, the proportion of people you meet that satisfy the listed criterion (e.g. 30% are intelligent enough). Think of each of these numbers as being the probability that a new person you meet will satisfy the listed criterion.

(c) Assume independence of the listed criteria. How do you get from the individual probabilities in (b) to the probability that a new person you meet will satisfy every criterion on your list? Make the calculation.

(d) Which criteria on your list, if any, are obviously not independent? Which of these are positively associated? Which are negatively associated?

(e) Assume you meet 300 new people a year. From the calculation in (c), how many years would you expect it will take to you, on average, to find a person meeting all your essential criteria?

(f) Are you depressed? Are some of your criteria not so essential after all? Which items on the list are making the search so difficult? How do the associations affect the answer? If you are still interested in this problem, read Barry Singer's chapter!

26. Consider the life table given in Table 5. It gives the numbers of males and females surviving at different ages per 100,000 born and can thus be interpreted as a way of presenting the probability that a randomly chosen baby will still be alive at each age.

Table 5 : Numbers Surviving per 100,000 Born[a]

Age (years)	Males	Females
0	100,000	100,000
1	98,826	99,035
5	98,565	98,827
10	98,414	98,692
15	98,216	98,569
20	97,473	98,304
25	96,594	97,998
30	95,851	97,684
35	95,180	97,317
40	94,407	96,792
60	82,690	89,099

[a]Source: NZ Life Tables, 1985-1987.

(**a**) Find the probability that: **(i)** a boy who survives the first year of life will reach 20;   **(ii)** a 20 year old woman will reach 60.

(**b**) By evaluating the probabilities of dying within a 5 year period given being alive at the start of the period, what is the second most dangerous 5-year period up to age 40 **(i)** for a male?   **(ii)** for a female? You may find the answer to (i) surprising. Can you think of an explanation?

(**c**) For a 20 year old man and a 20 year old woman, what is the probability that: **(i)** both will reach 60?   **(ii)** the woman will reach 60 but the man will die before 60?   **(iii)** neither will reach 60.

(**d**) What assumption did you have to make about the lifetimes of the two people in part (c) in order to do the calculations? Would this assumption hold if the man and woman were a married couple? If you think the assumption would fail, would you expect the association between lifetimes to be positive or negative? Why?

(**e**) Assume that 20 year old men and women exist in equal numbers. What is the probability that a random 20 year old person will live until 40?

Let $t_0 = 0, t_1, t_2, \ldots, t_n, \ldots$ be the ages in the life table at which survival probabilities are given. Population life tables are not, in fact, constructed by following a group of people and counting how many are still alive at age t_1, at age t_2 and so on. Let A_i represent the event "alive at age t_i".

(**f**) By using the generalized multiplication rule and considering the nature of the intersections, show that

$$\mathrm{pr}(A_n) = \mathrm{pr}(A_1)\mathrm{pr}(A_2|A_1)\mathrm{pr}(A_3|A_2)\ldots\mathrm{pr}(A_n|A_{n-1}).$$

The life table is constructed using this relationship and data collected at the current time.

(**g**) How would you estimate estimate an event like $\mathrm{pr}(A_i \,|\, A_{i-1})$, for example pr(still alive at age 25 | alive at age 20), on the basis of 1995 statistics?

(**h**) What assumption is being made by applying the answers in (a)(ii), (c) and (e) to people who are currently 20 years old? How valid do you think the assumption is?

27. The distribution of blood types for the New Zealand European population is as follows:

   40% type A,   9% type B,   49% type O,   2% type AB.

   Suppose that the blood types of European married couples are independent and that both the husband and the wife follow this distribution of blood type. In the following questions, we assume that the couple is randomly selected.

   (**a**) If the wife has type B blood, what is the probability that the husband has type B blood?

   (**b**) What is the probability that both the husband and wife have type B blood?

   (**c**) What is the probability that at least one member of the couple has type B blood?

   (**d**) What is the probability that both the husband and wife have the same type blood?

   (**e**) An individual with type B blood can safely receive transfusions only from persons with type B or type O blood. What is the probability that the husband of a woman with type B blood is an acceptable blood donor for her?

   (**f**) Construct a two-way table of probabilities for the various husband and wife blood-type combinations. Use the table to find your answers for (a) to (e).

28. In her 9 September 1990 "Ask Marilyn" column in *Parade Magazine*, Marilyn vos Savant (reported to be the holder of the world's highest I.Q.) posed a problem that attracted thousands of replies and lots of the controversy. The problem is as follows. "Suppose you're on a game show and given a choice of three doors. Behind one is a car; behind the others are goats. You pick door No. 1, and the host, who knows what is behind them, opens No. 3 which has a goat. He then asks if you want to pick No. 2. Should you switch?" Although it is not in the original statement of the problem it is clear from Marilyn's replies in subsequent columns that the Host will always open a door which does not belong to the player and has a goat behind it.[41] What is the probability of winning the car in a situation like this if your strategy is always to switch? never to switch?

29. The following, an historic problem called Bertrand's Box Problem, is taken from R.G. Seyman's discussion of Morgan et al. [1991]. A box contains three drawers, one containing two gold coins, one containing two silver

---

[41]It is possible, though more complicated, to solve the problem without knowing this – see Morgan et al. [1991].

coins, and one containing one gold and one silver coin. A drawer is chosen at random and a coin that is randomly selected from that drawer. If the selected coin turns out to be gold, what is the probability that the chosen drawer is the one with the two gold coins?[42]

30. Digital data is transmitted as a sequence of signals which represent 0's and 1's. Suppose that such data is being transmitted to a satellite and then relayed to distant site. Suppose that, due to electrical interference in the atmosphere, there is a 1 in 1,000 chance that a transmitted 0 will be reversed between the sender and satellite (i.e. distorted to the extent that the satellite's receiver interprets it as a 1) and a 2 in 1,000 chance that a transmitted 1 will be reversed. Suppose that 40% of transmitted digits are 0's.

(**a**) What is the probability that a transmitted digit is correctly received by the satellite?

(**b**) Assuming independence, what is the probability that all digits are received correctly **(i)** if 1,000 digits are transmitted? **(ii)** if 10,000 are transmitted?

(**c**) Suppose that between the satellite and the receiver, the chances of reversal are twice as large as they were between the sender and satellite. Assuming independence, what is the probability that a digit reaches the receiver as originally sent?

31. In reliability theory, components in a system are said to be *in parallel* if the system fails only if all the components in the system fail. For example, if we have an alarm warning system which gives warnings via a flashing light, a siren and a digital display, the warning system fails only if all three devices fail to alert people to the problem.

(**a**) Suppose that the light fails with probability 0.01, the siren with probability 0.02 and the display with probability 0.08. Assuming failures occur independently, what is the probability that the system fails?

(**b**) Particularly where components fail independently, putting parallel components into a system can make the system dramatically more reliable than any of its components, as in (a). However, the results are not as good as the independence calculations show if there are common causes of failure (resulting in positive association). Can you think of a possible common cause of failure for all three components above?

---

[42]Another version of this problem presents it as a card con trick. There are 3 cards with their two sides colored respectively as red/red, red/black and black/black. A card is chosen at random. The player can see the top side is red and has to guess the bottom side. Most people intuitively think that the bottom side is equally likely to be red or black. There is a potential for cheating gamblers out of money because the actual probability that the bottom side is red is quite different from 50%.

(**c**) Can you think of anything from your own experience where a subsystem of components act in parallel? (It need not be a mechanical or electronic system, it could be an administrative system.)

At the other extreme from a parallel subsystem, a set of components is said to be *in series* when the system fails if any one of the components fail. As a (simplified) example, suppose you plan to travel to Los Angeles for a meeting by driving to the airport, catching a flight to Los Angeles and picking up a hire car and driving to the site of the meeting. There are 4 "components" discussed here, the initial drive, the flight, picking up the hire car, and another drive (we could obviously break each of these down further) which are "in series". Suppose that you have estimated the probabilities of something going wrong with each of these components which would stop your getting to the meeting on time to be 0.02, 0.05, 0.08, and 0.03 respectively.

(**d**) Assuming independent causes of failure, what is the probability of failing to make the meeting on ? [Hint: It is easier to work with succeeding than failing for series systems.]

Note from your answer how, in contrast to a parallel system, the series system is much less reliable than any of its components.

(**e**) Can you think of any common causes of failure here?

(**f**) What are some subsystems of series components that you use or have experienced?

Suppose that we complicate the system a little by allowing "taking a taxicab" as an alternative to hiring and driving a car for getting from the airport to the final destination and that this alternative has probability 0.04 of failure.

(**g**) In terms of the failure of each "component" and its complement (success), define the possible outcomes in the event "making it to the meeting"

(**h**) Again assuming that all failures are independent, what is the probability of failing to make the meeting now?

*32. A large study of college students by Professor Stanley Coren of the University of British Columbia and Dr Diane Halpern of San Bernardino State University showed that left-handed people had an 89% greater risk of having a car accident than right-handed people (*NZ Herald*, 5 January 1991). Left-handed people make up 10% of drivers. What percentage of people having car accidents are left-handed?

[Hint: Work with p where p denotes the probability that a right handed person has a car accident. You do not actually need to know p to solve the problem, but if you are having trouble working with an unknown p, assume a value for it and see if you can see why your answer does not depend upon the value assumed.]