Chapter 8

Discrete Probability Models

8.1 Introduction

The link between probability and statistics arises because in order to see, for example, how strong the evidence is in some data, we need to consider the probabilities concerned with how we came to observe the data. In this chapter, we describe some standard probability models which are often used used with data from various sources, such as market research. However, before we describe these in detail, we need to establish some ground rules for counting.

8.2 Permutations and Combinations

8.2.1 Numbers of sequences

Imagine that your cash point card has just been stolen. What is the probability of the thief guessing your 4 digit PIN in one go? To answer this question, we need to know how many different 4 digit PINs there are. We are also assuming that the thief chooses in such a way that all possibilities are equally likely. With this assumption the probability of a correct guess (in one go) is

$$\begin{aligned} P(\text{Guess correctly}) &= \tfrac{\text{number of correct PINs}}{\text{number of possible 4 digit PINs}} \\ &= \tfrac{1}{\text{number of possible 4 digit PINs}}. \end{aligned}$$

There is, of course, only one correct PIN. The number of possible 4 digit PINs is calculated as follows. There are 10 choices for the first digit, another 10 choices for the second digit, and so on. Therefore the number of possible choices is

$$10 \times 10 \times 10 \times 10 = 10,000.$$

So the probability of a correct guess is

$$\begin{aligned} P(\text{Guess correctly}) &= \tfrac{1}{10\times10\times10\times10} \\ &= \tfrac{1}{10,000} \\ &= 0.0001. \end{aligned}$$

8.2.2 Permutations

The calculation of the card-thief's correct guess of a PIN changes if the thief knows that your PIN uses 4 different digits. Now the number of possible PINs is smaller. To find this number we need to work out how many ways there are to arrange 4 digits out of a list of 10.

In more general terms, we need to know how many different ways there are of arranging r objects from a list of n objects. The best way of thinking about this is to consider the choice of each item as a different experiment. The first experiment has n possible outcomes. The second experiment only has $n-1$ possible outcomes, as one has already been selected. The third experiment has $n-2$ possible outcomes and so on, until the rth experiment which has $n-r+1$ possible outcomes. Therefore the number of possible selections is

$$n \times (n-1) \times (n-2) \times \ldots \times (n-r+1) = \frac{n!}{(n-r)!},$$

where

$$n! = n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 3 \times 2 \times 1$$

and is know as n *factorial* – it can be found on most calculators. In fact, the right–hand–side of the first equation above is a commonly encountered expression in counting calculations and has its own notation. The number of ordered ways of selecting r objects out of n is denoted ${}^n\mathrm{P}_r$, where

$${}^n\mathrm{P}_r = \frac{n!}{(n-r)!}.$$

and this can also be found on most calculators (sometimes it looks like $n\mathrm{P}r$). We refer to ${}^n\mathrm{P}_r$ as the number of *permutations* of r objects out of n.

Returning to the example in which the thief is trying to guess your 4–digit PIN, if the thief knows that the PIN contains no repeated digits then the number of possible PINs is

$${}^{10}\mathrm{P}_4 = 5040 \ (= 10 \times 9 \times 8 \times 7)$$

so, assuming that each is equally likely to be guessed, the probability of a correct guess is

$$P(\text{Guess correctly}) = \frac{1}{5040} = 0.0001984.$$

This illustrates how important it is to keep secret all information about your PIN. These probability calculations show that even knowing whether or not you repeat digits in your PIN is informative for a thief – it reduces the number of possible PINs by a factor of around 2.

8.2.3 Combinations

We now have a way of counting permutations, but often when selecting objects, all that matters is *which* objects were selected, not the order in which they were selected. Suppose that we have a collection of n objects and that we wish to make r selections from this list of objects, where the order does not matter. An unordered selection such as this is referred to as a *combination*. How many ways can this be done? Notice that this is equivalent to asking how many different ways are there of choosing r objects from n objects.

For example, a company has 20 retail outlets. It is decided to try a sales promotion at 4 of these outlets. How many selections of 4 can be chosen? It may be important to know this when we come to look at the results of the trial.

This calculation is very similar to that of permutations except that the ordering of objects no longer matters. For example, if we select two objects from three objects A, B and C, there are ${}^3\mathrm{P}_2 = 6$ ways of doing this:

$$A,\ B \qquad A,\ C \qquad B,\ A \qquad B,\ C \qquad C,\ A \qquad C,\ B.$$

However, if we are not interested in the ordering, just in whether A, B or C are chosen, then A, B is the same as B, A *etc.* and so the number of selections is just 3:

$$A,\ B \qquad A,\ C \qquad B,\ C.$$

In general, the number of *combinations* of r objects from n objects is

$${}^n\mathrm{C}_r = \frac{n!}{r!(n-r)!}.$$

There are other commonly used notations for this quantity: C_r^n, $\binom{n}{r}$ and often $n\mathrm{C}r$ on calculators. These numbers are known as the *binomial coefficients*.

Now we can see that the number of ways to select 4 retail outlets out of 20 is

$${}^{20}\mathrm{C}_4 = \frac{20!}{4!16!} = 4845.$$

To see how combinations can be used to calculate probabilities, we will look at the UK National Lottery example we touched on in last week's lecture (Chapter 7). In this lottery, there are 49 numbered balls, and six of these are selected at random. A seventh ball is also selected, but this is only relevant if you get exactly five numbers correct. The player selects six numbers before the draw is made, and after the draw, counts how many numbers are in common with those drawn. Players win a prize if they select at least three of the balls drawn. The order in which the balls are drawn in is irrelevant.

Let's consider the probability of winning the jackpot (i.e. you correctly match all six balls). How many ways can 6 balls be chosen out of 49? One option is $\{1, 2, 3, 4, 5, 6\}$; another is $\{1, 2, 3, 4, 5, 7\}$ ... in fact, there are

$${}^{49}\mathrm{C}_6 = 13,983,816$$

different ways 6 balls can be selected out of a possible 49. Now out of these 13,983,816 different combinations, how many combinations match the drawn balls correctly? Only one! There is only one set of six numbers that wins the jackpot! So the probability of winning the jackpot is just one in 13,983,816, i.e.

$$P(\text{match exactly 6 correct numbers}) = \frac{1}{13,983,816},$$

or just over a one in fourteen million chance! The other probabilities used last in last week's lecture to calculate the Expected Monetary Value for the lottery can be found using similar arguments.

8.3 Probability Distributions

8.3.1 Introduction

In Chapter 1 we saw how surveys can be used to get information on population quantities. For example, we might want to know voting intentions within the UK just before a General Election. Why does this involve *random* variables? In most cases, it is not possible to measure the variables on every member of the population and so some sampling scheme is used. This means that there is uncertainty in our conclusions. For example, if the true proportion of Labour voters were 40%, in a survey of 1,000 voters, it would be possible to get 400 Labour voters, but it would also be possible to get 350 Labour voters or 430 Labour voters. The fact that we have only a sample of voters introduces uncertainty into our conclusions about voting intentions in the population as a whole. Sometimes experiments themselves have inherent variability, for example, the toss of a coin. If the coin were tossed 1000 times and heads occurred only 400 times, would it be fair to conclude that the coin was a biased coin? The subject of Statistics has been developed to understand such variability and, in particular, how to draw correct conclusions from data which are subject to experimental and sampling variability.

Before we can make inferences about populations, we need a language to describe the uncertainty we find when taking samples from populations. First, we represent a random variable by X (capital X) and the probability that it takes a certain value x (small x) as $P(X = x)$.

The *probability distribution* of a discrete random variable X is the list of all possible values X can take and the probabilities associated with them. For example, if the random variable X is the outcome of a roll of a die then the probability distribution for X is:

x	$P(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
sum	1

In the die–rolling example, we used the *classical* interpretation of probability to obtain the probability distribution for X, the outcome of a roll on the die. Consider the following *frequentist* example.

Let X be the number of cars observed in half–hour periods passing the junction of two roads. In a five hour period, the following observations on X were made:

$$2 \quad 3 \quad 2 \quad 5 \quad 5 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7$$

Obtain the probability distribution of X.

We can calculate the following probabilities:

$$\begin{aligned} P(X = 0) &= \\ P(X = 1) &= \\ P(X = 2) &= \\ P(X = 3) &= \\ P(X = 4) &= \\ P(X = 5) &= \\ P(X = 6) &= \\ P(X = 7) &= \\ P(X > 7) &= \end{aligned}$$

Thus would give:

x	$P(X = x)$
< 2	
2	
3	
4	
5	
6	
7	
> 7	
sum	

Does this make sense? Is it *impossible* to see, for example, 8 cars at this junction? Do we know, with certainty, that we will *never* see 10 cars pass this junction, even in rush hour? Do we know that we will *never* see fewer than two cars pass this junction, even at three o'clock in the morning? We often need to make predictions based on the limited information provided by a sample as small as this, but these predictions need to be sensible. This is one situation where *probability models* are useful. In this Chapter, we consider two such models for discrete data.

8.3.2 The Binomial Distribution

In many surveys and experiments data is collected in the form of counts. For example, the number of people in the survey who bought a CD in the past month, the number of people who said they would vote Labour at the next election, the number of defective items in a sample taken from a production line, and so on. All these variables have common features:

1. Each person/item has only two possible (exclusive) responses (Yes/No, Defective/Not defective etc)
   – this is referred to as a *trial* which results in a *success* or *failure*

2. The survey/experiment takes the form of a random sample
   – the responses are independent.

Further suppose that the true probability of a success in the population is p (in which case the probability of a failure is $1 - p$). We are interested in the random variable X, the total number of successes out of n trials.

Example
suppose we are interested in the number of sixes we get from 4 rolls of a dice. Each roll of the dice is a trial which gives a "six" (success, or s) or "not a six" (failure, or f). The probability of a success is $p = P(\text{six}) = 1/6$. We have $n = 4$ independent trials (rolls of the dice). Let X be the number of sixes obtained. We can now obtain the full probability distribution of X.

For example, suppose we want to work out the probability of obtaining four sixes (four "successes" – i.e. $ssss$ – or $P(X = 4)$). Since the rolls of the die can be considered independent, we get:

$$\begin{aligned} P(ssss) &= P(s) \times P(s) \times P(s) \times P(s) \\ &= \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \\ &= \left(\frac{1}{6}\right)^4 \end{aligned}$$

That one's easy! What about the probability that we get three sixes – i.e. $P(X = 3)$? This one's a bit more tricky, because that means we need three s's and one f – i.e. three sixes and one "not six" – but the "not six" could appear on the first roll, or the second roll, or the third, or the fourth! For example, for $P(X = 3)$, we could have:

$$P(fsss) = \frac{5}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^3 \times \frac{5}{6},$$

or we could have:

$$P(sfss) = \frac{1}{6} \times \frac{5}{6} \times \frac{1}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^3 \times \frac{5}{6},$$

or maybe:

$$P(ssfs) = \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} \times \frac{1}{6} = \left(\frac{1}{6}\right)^3 \times \frac{5}{6},$$

or even:

$$P(sssf) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} = \left(\frac{1}{6}\right)^3 \times \frac{5}{6}.$$

Can you see that we therefore get:

$$P(X = 3) = 4 \times \left(\frac{1}{6}\right)^3 \times \frac{5}{6}.$$

Thinking about it, there are actually sixteen possible outcomes for the four rolls of the die:

	Outcome	**Probability**
1	$ssss$	$\left(\frac{1}{6}\right)^4$
2	$fsss$	$\left(\frac{1}{6}\right)^3 \times \frac{5}{6}$
3	$sfss$	$\left(\frac{1}{6}\right)^3 \times \frac{5}{6}$
4	$ssfs$	$\left(\frac{1}{6}\right)^3 \times \frac{5}{6}$
5	$sssf$	$\left(\frac{1}{6}\right)^3 \times \frac{5}{6}$
6	$ssff$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
7	$ffss$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
8	$sfsf$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
9	$fsfs$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
10	$sffs$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
11	$fssf$	$\left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2$
12	$sfff$	$\frac{1}{6} \times \left(\frac{5}{6}\right)^3$
13	$fsff$	$\frac{1}{6} \times \left(\frac{5}{6}\right)^3$
14	$ffsf$	$\frac{1}{6} \times \left(\frac{5}{6}\right)^3$
15	$fffs$	$\frac{1}{6} \times \left(\frac{5}{6}\right)^3$
16	$ffff$	$\left(\frac{5}{6}\right)^4$

So we get:

$$\begin{aligned}
P(X = 4) &= \left(\frac{1}{6}\right)^4 = 0.0008 \\
P(X = 3) &= 4 \times \left(\frac{1}{6}\right)^3 \times \frac{5}{6} = 0.0153 \\
P(X = 2) &= 6 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2 = 0.1158 \\
P(X = 1) &= 4 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^3 = 0.3858 \qquad \text{and} \\
P(X = 0) &= \left(\frac{5}{6}\right)^4 = 0.4823,
\end{aligned}$$

and so the full probability distribution for X is:

x	0	1	2	3	4
$P(X=x)$	0.4823	0.3858	0.1158	0.0153	0.0008

Now this is a bit long–winded... and that was just for four rolls of the die! Note there were *four* different ways we could get $X=3$, and *six* different ways we could get $X=2$, but these numbers would increase substantially if we considered more than four rolls of the die. We would like a more concise way of working these probabilities out without having to list all the possible outcomes as we did above. And luckily, there's a formula that does just that!

We can calculate probabilities (providing the above assumptions are reasonable), using the following formula:

$$P(X=r) = {}^n\mathrm{C}_r\, p^r(1-p)^{n-r}, \quad r=0,1,\ldots,n.$$

These probabilities describe how likely we are to get r out of n successes from independent trials, each with success probability p. Note that any number raised to the power zero is one, for example, $0.3^0=1$ and $0.654^0=1$.

This distribution is known as the *binomial distribution* with index n and probability p. We write this as $X \sim Bin(n,p)$, which reads as "X has a binomial distribution with index n and probability p". Here, n and p are known as the "parameters" of the binomial distribution.

For example, in the die example, we know $n=4$ and $p=P(\text{six})=1/6$. Each roll of the dice is a trial which gives a "six" (success) or "not a six" (failure). The probability of a success is $p=P(\text{six})=1/6$. We have $n=4$ independent trials (rolls of the dice). If X is the number of sixes obtained then $X \sim Bin(4,1/6)$, and so

$$\begin{aligned} P(X=0) &= {}^4\mathrm{C}_0 \left(\frac{1}{6}\right)^0 \left(1-\frac{1}{6}\right)^4 \\ &= 1 \times 1 \times 0.4823 \\ &= 0.4823 \end{aligned}$$

$$\begin{aligned} P(X=1) &= {}^4\mathrm{C}_1 \left(\frac{1}{6}\right)^1 \left(1-\frac{1}{6}\right)^3 \\ &= 4 \times 0.1667 \times 0.5787 \\ &= 0.3858 \end{aligned}$$

$$\begin{aligned} P(X=2) &= {}^4\mathrm{C}_2 \left(\frac{1}{6}\right)^2 \left(1-\frac{1}{6}\right)^2 \\ &= 6 \times 0.0278 \times 0.6944 \\ &= 0.1158 \end{aligned}$$

$$\begin{aligned}P(X=3) &= {}^4\mathrm{C}_3\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^1\\ &= 4\times 0.0046\times 0.8333\\ &= 0.0153\end{aligned}$$

$$\begin{aligned}P(X=4) &= {}^4\mathrm{C}_4\left(\frac{1}{6}\right)^4\left(1-\frac{1}{6}\right)^0\\ &= 1\times 0.0008\times 1\\ &= 0.0008.\end{aligned}$$

This probability distribution shows that most of the time we would get either 0 or 1 successes and, for example, 4 successes would be quite rare.

Let's see how close these "theoretical" probabilities are to some "observed" values obtained by actually rolling a dice four times and counting the number of sixes we get. Actually, we'll not roll a dice, but will use Minitab instead!

No. of sixes	Binomial probability	Observed probability
0	0.4823	
1	0.3858	
2	0.1158	
3	0.0153	
4	0.0008	
Sum		

Another example
A salesperson has a 50% chance of making a sale on a customer visit and she arranges 6 visits in a day. What are the probabilities of her making 0,1,2,3,4,5 and 6 sales? Let X denote the number of sales. Assuming the visits result in sales independently, $X \sim Bin(6, 0.5)$ and

No. of sales r	Probability $P(X=r)$	Cumulative Probability $P(X \leq r)$
0	0.015625	0.015625
1	0.093750	0.109375
2	0.234375	0.343750
3	0.312500	0.656250
4	0.234375	0.890625
5	0.093750	0.984375
6	0.015625	1.000000
sum	1.000000	

The formula for binomial probabilities enables us to calculate values for $P(X = r)$. From these, it is straightforward to calculate cumulative probabilities such as the probability of making no more than 2 sales:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.015625 + 0.09375 + 0.234375 = 0.34375. \end{aligned}$$

These cumulative probabilities are also useful in calculating probabilities such as that of making more than 1 sale:

$$P(X > 1) = 1 - P(X \leq 1) = 1 - 0.109375 = 0.890625.$$

If we have the probability distribution for X rather than the raw observations, we denote the mean for X not by $\bar{x}$ but with $E(X)$ (which reads as "the expectation of X"), and the variance with $Var(X)$.

If X is a random variable with a binomial $Bin(n, p)$ distribution then its mean and variance are

$$E(X) = np, \qquad Var(X) = np(1 - p).$$

For example, if $X \sim Bin(6, 0.5)$ then

$$E(X) = np = 6 \times 0.5 = 3$$

and

$$Var(X) = np(1 - p) = 6 \times 0.5 \times 0.5 = 1.5.$$

Also

$$SD(X) = \sqrt{Var(X)} = \sqrt{1.5} = 1.225.$$

8.3.3 The Poisson Distribution

The *Poisson distribution* is a very important discrete probability distribution which arises in many different contexts. Typically, Poisson random quantities are used in place of binomial random quantities in situations where n is large, p is small, and both np and $n(1-p)$ exceed 5. In general, it is used to model data which are counts of (random) events in a certain area or time interval, without a known fixed upper limit but with a known *rate* of occurrence.

For example, consider the number of calls made in a 1 minute interval to an Internet service provider (ISP). The ISP has thousands of subscribers, but each one will call with a very small probability. If the ISP knows that on average 5 calls will be made in the interval, the actual number of calls will be a Poisson random variable, with mean 5.

If X is a random variable with a Poisson distribution with parameter λ (Greek lower case *lambda*) then the probability it takes different values is

$$P(X = r) = \frac{\lambda^r \, e^{-\lambda}}{r!}, \quad r = 0, 1, 2, \ldots \; .$$

We write this as $X \sim Po(\lambda)$. The parameter λ has a very simple interpretation as the rate at which events occur. The distribution has mean and variance

$$E(X) = \lambda, \qquad Var(X) = \lambda.$$

Thus, when approximating binomial probabilities by Poisson probabilities, we match the means of the distributions: $\lambda = np$.

Returning to the ISP example, suppose we want to know the probabilities of different numbers of calls made to the ISP. Let X be the number of calls made in a minute. Then $X \sim P(5)$ and, for example, the probability of receiving 4 calls is

$$P(X = 4) = \frac{5^4 e^{-5}}{4!} = 0.1755.$$

We can use the formula for Poisson probabilities to calculate the probability of all possible outcomes:

r	Probability $P(X = r)$	Cumulative Probability $P(X \leq r)$
0	0.0067	0.0067
1	0.0337	0.0404
2	0.0843	0.1247
3	0.1403	0.2650
4	0.1755	0.4405
5	0.1755	0.6160
6	0.1462	0.7622
7	0.1044	0.8666
8	0.0653	0.9319
$\vdots$	$\vdots$	$\vdots$
sum	1.000000	

Therefore the probability of receiving between 2 and 8 calls is

$$P(2 \leq X \leq 8) = P(X \leq 8) - P(X \leq 1) = 0.9319 - 0.0404 = 0.8915$$

and so is very likely. Probability calculations such as this enable ISPs to calibrate the likely demand for their service and hence the resources they need to provide the service. Using such a model we can also account for "extreme" situations. For example, suppose that, for this ISP, we observed the following number of calls per minute over a five minute period: 6, 3, 5, 4, 6. Using simple frequentist reasoning, we would have

$$P(7 \text{ calls made}) \quad = \quad \frac{0}{5} = 0,$$

i.e. we will *never* observe seven calls in any one minute period! However, using the Poisson model, we have

$$P(X = 7) \quad = \quad 0.1044,$$

which is probably more realistic. You can see from the previous table of probabilities that, although when r is large the associated probabilities are small, at least they are accounted for and are, more realistically, non–zero.

8.4 Exercises

1. Consider a lottery that is slightly different to the National Lottery in that there are 48 balls instead of 49. What is the probability of winning the jackpot in this lottery?

2. A market survey has identified 10 desirable features for a new product. However, due to cost constraints, only four of these features can be included. If the features are selected randomly, what is the probability that your four favourites are chosen in your preferred ordering?

3. If you dial a random 7 digit local telephone number, what is the probability you dial your own number?

4. An operator at a call centre has 20 calls to make in an hour. History suggests that they will be answered 60% of the time. Let X be the number of answered calls in an hour.

   (a) What probability distribution does X have?

   (b) What is the mean and standard deviation of X?

   (c) Calculate the probability of getting a response exactly 9 times.

   (d) Calculate the probability of getting less than 2 responses.

5. Calls are received at a telephone exchange at an average rate of 10 per minute. Let X be the number of calls received in one minute.

   (a) What probability distribution does X have?

   (b) What is the mean and standard deviation of X?

   (c) Calculate the probability that there are 12 calls in one minute.

   (d) Calculate the probability there are no more than 2 calls in a minute.