MAS187/AEF258

University of Newcastle upon Tyne

2005-6

Contents

1 Collecting and Presenting Data **5**

1.1 Introduction 5

1.1.1 Examples 5

1.1.2 Definitions 5

1.1.3 Surveys 6

1.2 Sampling 7

1.2.1 Simple Random Sampling 7

1.2.2 Stratified Sampling 8

1.2.3 Systematic Sampling 8

1.2.4 Multi-stage Sampling 8

1.2.5 Cluster Sampling 9

1.2.6 Judgemental sampling 9

1.2.7 Accessibility sampling 9

1.2.8 Quota Sampling 9

1.2.9 Sample Size 10

1.3 Frequency Tables 10

1.3.1 Frequency Tables 10

1.3.2 Continuous Data Frequency Tables 12

1.4 Exercises 1 14

2 Graphical methods for presenting data **15**

2.1 Introduction 15

2.2 Stem and Leaf plots 15

2.2.1 Using Minitab 17
2.3 Bar Charts 19
2.4 Multiple Bar Charts 22
2.5 Histograms 24
2.6 Exercises 2 30

3 More graphical methods for presenting data **31**
3.1 Introduction 31
3.2 Percentage Relative Frequency Histograms 31
3.3 Relative Frequency Polygons 34
3.4 Cumulative Frequency Polygons (Ogive) 38
3.5 Pie Charts 41
3.6 Time Series Plots 43
3.7 Scatter Plots 46
3.8 Exercises 3 48

4 Numerical summaries for data **51**
4.1 Introduction 51
4.2 Mathematical notation 51
4.3 Measures of Location 52
4.3.1 The Arithmetic Mean 52
4.3.2 The Median 54
4.3.3 The Mode 55
4.4 Measures of Spread 56
4.4.1 The Range 56
4.4.2 The Inter-Quartile Range 56
4.4.3 The Sample Variance and Standard Deviation 57
4.5 Summary statistics in **MINITAB** 59
4.6 Box and Whisker Plots 60
4.7 Exercises 4 63

5 Introduction to Probability **64**

5.1 Introduction . . . 64

5.1.1 Definitions . . . 64

5.2 How do we measure Probability? . . . 65

5.2.1 Classical . . . 65

5.2.2 Frequentist . . . 66

5.2.3 Subjective/Bayesian . . . 66

5.3 Laws of Probability . . . 67

5.3.1 Multiplication Law . . . 67

5.3.2 Addition Law . . . 67

5.3.3 Example . . . 68

5.4 Exercises 5 . . . 69

6 Decision Making using Probability **71**

6.1 Conditional Probability . . . 71

6.2 Multiplication of probabilities . . . 72

6.3 Tree Diagrams . . . 74

6.4 Expected Monetary Value and Probability Trees . . . 76

6.5 Exercises 6 . . . 78

7 Discrete Probability Models **80**

7.1 Introduction . . . 80

7.2 Permutations and Combinations . . . 80

7.2.1 Permutations . . . 80

7.2.2 Combinations . . . 82

7.3 Probability Distributions . . . 84

7.3.1 Expectation and the population mean . . . 85

7.3.2 Population variance and standard deviation . . . 85

7.4 The Binomial Distribution . . . 86

7.5 The Poisson Distribution . . . 89

7.6 Exercises 7 . . . 92

8 Continuous Probability Models **93**

8.1 Introduction 93

8.2 The Uniform Distribution 94

8.2.1 Mean and Variance 96

8.3 The Exponential Distribution 96

8.3.1 Mean and Variance 97

8.4 The Normal Distribution 98

8.4.1 Notation 99

8.4.2 Probability calculations and the standard normal distribution 99

8.5 Exercises 8 104

Chapter 6

Decision Making using Probability

In this chapter, we look at more complicated notions of probability and how we can use probability in order to aid in management decision making.

6.1 Conditional Probability

So far we have only considered probabilities of single events or of several independent events, like two rolls of a die. However in reality many events are related. For example the probability of it raining in 5 minutes time is dependent on whether or not it is raining now.

We need a mathematical notation to capture how the probability of one event depends on other events taking place. We do this as follows. Consider two events A and B. We write

$$P(A|B)$$

for the probability of A given B has already happened. We describe $P(A|B)$ as the conditional probability of A given B. For example, the probability of it raining in 5 minutes time given that it is raining now would be

$$P(\text{Rain in 5 minutes}|\text{Raining now}).$$

Utility companies need to be able to forecast periods of high demand. They describe their forecasts in terms of probabilities. Gas and electricity suppliers might relate them to air temperature. For example,

$$P(\text{High demand}|\text{air temperature is below normal}) = 0.6$$
$$P(\text{High demand}|\text{air temperature is normal}) = 0.2$$
$$P(\text{High demand}|\text{air temperature is above normal}) = 0.05.$$

We can calculate these conditional probabilities using the formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)},$$

that is, in terms of the probability of both events occurring, $P(A \text{ and } B)$, and the probability of the event that has already taken place, $P(B)$.

To see how this formula works, let's consider a simple example based on the class of students in Exercises 5.

Student Number	Sex	Height (m)	Weight (kg)	Shoe Size	Student Number	Sex	Height (m)	Weight (kg)	Shoe Size
1	M	1.91	70	11.0	10	M	1.78	76	8.5
2	F	1.73	89	6.5	11	M	1.88	64	9.0
3	M	1.73	73	7.0	12	M	1.88	83	9.0
4	M	1.63	54	8.0	13	M	1.70	55	8.0
5	F	1.73	58	6.5	14	M	1.76	57	8.0
6	M	1.70	60	8.0	15	M	1.78	60	8.0
7	M	1.82	76	10.0	16	F	1.52	45	3.5
8	M	1.67	54	7.5	17	M	1.80	67	7.5
9	F	1.55	47	4.0	18	M	1.92	83	12.0

Suppose we want the probability that a student chosen at random from this class will be female given that the student's shoe size is less than 8. We could simply find the proportion of students with shoe sizes less than 8 who are female. There are 7 students with shoe sizes less than 8 and 4 of these are female. So

$$P(\text{Female}|\text{Shoe size less than 8}) = \frac{4}{7}.$$

This probability can also be calculated using the above formula as follows:

$$\begin{aligned} P(\text{Shoe size less than 8}) &= \frac{7}{18} \\ P(\text{Shoe size less than 8 and female}) &= \frac{4}{18} \end{aligned}$$

and so

$$P(\text{Female}|\text{Shoe size less than 8}) = \frac{P(\text{Shoe size less than 8 and female})}{P(\text{Shoe size less than 8})} = \frac{4/18}{7/18} = \frac{4}{7}.$$

6.2 Multiplication of probabilities

We saw in Chapter 5 that, if two events A and B are independent, then $P(A \text{ and} B) = P(A)P(B)$. Now we know that

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)},$$

we can easily see that

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

Of course it is also true that $P(A \text{ and } B) = P(A)P(B|A)$.

For example, consider a student chosen at random from the example class. Let F be the event "the student is female" and S be the event "the student's weight is less than 60kg." Then the probability that the student is female and has a weight less than 60kg is

$$\begin{aligned} P(F \text{and} S) &= P(S)P(F|S) = \frac{7}{18} \times \frac{3}{7} = \frac{3}{18} \\ &= P(F)P(S|F) = \frac{4}{18} \times \frac{3}{4} = \frac{3}{18} \end{aligned}$$

Notice that, if M is the event "the student is male," then $P(S|M) = 4/14 = 0.286$ and this is not equal to $P(S|F) = 3/4 = 0.75$. So the probability of the student having a weight less than 60kg depends on the student's sex, that is whether the student is female or male. The events S and F are not independent. Similarly $P(F|S) = 3/7 = 0.429$ while $P(F|L) = 1/11 = 0.091$, where L is the event "the students's weight is not less than 60kg." So, knowing whether or not a student's weight is less than 60kg gives us information about whether the student is likely to be male or female.

Let $\bar{B}$ be the event "not B." So, for example $\bar{F} = M$. Then we say that tow events A and B are independent if $P(A|B) = P(A|\bar{B}) = P(A)$. It is easy to show that this is equivalent to $P(B|A) = P(B|\bar{A}) = P(B)$. If A and B are independent then $P(A \text{and} B) = P(A)P(B)$.

For example, consider the following probabilities for customers at a cafe who can choose eiether icecream or treacle sponge and custard.

	Icecream	Treacle sponge
Male	0.250	0.150
Female	0.375	0.225

We see that $P(\text{male}) = 0.250+0.150 = 0.4$ and $P(\text{female}) = 0.375+0.225 = 0.6 = 1-P(\text{male})$. Now

$$P(\text{Icecream}|\text{Male}) = \frac{0.250}{0.4} = 0.625$$

and

$$P(\text{Icecream}|\text{Female}) = \frac{0.375}{0.6} = 0.625$$

so Icecream and Male are independent events. In fact the variables Sex and Dessert-choice are independent in this example. So the probability that a customer is male and chooses icecream is just $P(\text{Male})P(\text{Icecream}) = 0.4 \times 0.625 = 0.25$. (The probability of icecream is just $0.250 + 0.375 = 0.625$).

Another example relates to the age and sex distribution of purchasers of CD singles at an outlet:

	< 30	$30 - 50$	$50+$
Male	0.275	0.125	0.025
Female	0.325	0.175	0.075

From this table, we can calculate

$$\begin{aligned} P(\text{Male}) &= P(\text{Male and } < 30) + P(\text{Male and } 30 - 50) + P(\text{Male and } 50+) \\ &= 0.275 + 0.125 + 0.025 = 0.425 \end{aligned}$$

and

$$\begin{aligned}P(\text{Female}) &= P(\text{Female and } <30) + P(\text{Female and } 30-50) + P(\text{Female and } 50+)\\ &= 0.325 + 0.175 + 0.075 = 0.575.\end{aligned}$$

Also, the age distribution of the customers is

$$\begin{aligned}P(<30) &= P(\text{Male and } <30) + P(\text{Female and } <30) = 0.275 + 0.325 = 0.6\\ P(30-50) &= P(\text{Male and } 30-50) + P(\text{Female and } 30-50) = 0.125 + 0.175 = 0.3\\ P(50+) &= P(\text{Male and } 50+) + P(\text{Female and } 50+) = 0.025 + 0.075 = 0.1.\end{aligned}$$

Using this information we can calculate various probabilities such as:

$$\begin{aligned}P(\text{Male}|30-50) &= \frac{P(\text{Male and } 30-50)}{P(30-50)} = \frac{0.125}{0.3} = 0.4167\\ P(\text{Female}|30-50) &= 1 - P(\text{Male}|30-50) = 1 - 0.4167 = 0.5833\end{aligned}$$

and

$$\begin{aligned}P(<30|\text{Male}) &= \frac{P(\text{Male and } <30)}{P(\text{Male})} = \frac{0.275}{0.425} = 0.6471\\ P(30-50|\text{Male}) &= \frac{P(\text{Male and } 30-50)}{P(\text{Male})} = \frac{0.125}{0.425} = 0.2941\\ P(50+|\text{Male}) &= 1 - P(<30|\text{Male}) - P(30-50|\text{Male}) = 1 - 0.6471 - 0.2941 = 0.0588.\end{aligned}$$

6.3 Tree Diagrams

Tree diagrams or probability trees are simple clear ways of presenting probabilistic information. Let us first consider a simple example in which a die is rolled twice. Suppose we are interested in the probability that we score a six on both rolls. This probability can be calculated as

$$\begin{aligned}P(\text{Six and Six}) &= P(\text{Six on 1st throw}) \times P(\text{Six on 2nd throw}|\text{Six on 1st throw})\\ &= \frac{1}{6} \times \frac{1}{6}\\ &= \frac{1}{36}.\end{aligned}$$

This example can be represented as a tree diagram in which experiments are represented by circles (called *nodes*) and the outcomes of the experiments as *branches*. The branches are annotated by the probability of the particular outcome.

Here the probability of a six followed by a six is found by tracing the branch corresponding to this outcome through the tree. Note that the ends of the branches of the tree are usually known as *terminal nodes*.

Consider a more complicated example. A machine is used to produce components. Each time it produces a component there is a chance that the component will be defective. When the machine is working correctly the probability that a component is defective is 0.05. Sometimes, though, the machine requires adjustment and, when this is the case, the probability that a component is defective is 0.2. At the time in question there is a probability of 0.1 that the machine requires adjustment. Components produced by the machine are tested and either accepted or rejected. A component which is not defective is accepted with probability 0.97 and (falsely) rejected with probability 0.03. A defective component is (falsely) accepted with probability 0.15 and rejected with probability 0.85.

We can calculate various probabilities. For example:

$$
\begin{aligned}
P(\text{accepted}) &= 0.82935 + 0.00675 + 0.07760 + 0.00300 = 0.9167 \\
P(\text{defective}) &= (0.9 \times 0.05) + (0.1 \times 0.2) = 0.045 + 0.02 = 0.065 \\
P(\text{defective and accepted}) &= 0.00675 + 0.00300 = 0.00975 \\
P(\text{accepted} \mid \text{defective}) &= \frac{0.00975}{0.065} = 0.15 \\
P(\text{defective} \mid \text{accepted}) &= \frac{0.00975}{0.9167} = 0.010636 \\
P(\text{machine OK and accepted}) &= 0.82935 + 0.00675 = 0.8361 \\
P(\text{machine OK} \mid \text{accepted}) &= \frac{0.8361}{0.9167} = 0.9121 \\
P(\text{machine OK and rejected}) &= 0.02565 + 0.03825 = 0.0639 \\
P(\text{rejected}) &= 1 - P(\text{accepted}) = 0.0833 \\
P(\text{machine OK} \mid \text{rejected}) &= \frac{0.0639}{0.0833} = 0.7671
\end{aligned}
$$

6.4 Expected Monetary Value and Probability Trees

Probability trees can be used to see the effect of making particular decisions in the face of uncertainty. This is achieved by weighting the probability of different outcomes by their value. Often this value is financial. The *Expected Monetary Value (EMV)* of a single event is simply the probability of that event multiplied by the monetary value of that outcome. For example, if you would win £5 if you pulled an ace from a pack of cards, the EMV would be

$$EMV(Ace) = \frac{1}{13} \times 5 = 0.38.$$

In other words, if you repeated this bet a large number of times, overall you would come out an average of 38 pence better off per bet. Therefore you would want to pay no more than $38p$ for such a bet.

Consider another bet. When rolling a die, if it's a six you have to pay £5 but if it's any other number you receive £2.50. Would you take on this bet?

Probability	**Financial outcome**
$P(6) = 1/6$	-£5
$P(\text{Not a } 6) = 5/6$	£2.50

Therefore

$$EMV(\text{Six}) = \frac{1}{6} \times -5.00 = -0.833$$
$$EMV(\text{Not a Six}) = \frac{5}{6} \times 2.50 = 2.0833$$

and hence the expected monetary value of the bet is

$$EMV(\text{Bet}) = -0.833 + 2.083 = 1.25.$$

Therefore, in the long run, this would be a bet to take on as it has a positive expected monetary value.

In general, the expected monetary value of a project or bet is given by the formula

$$EMV = \sum P(\text{Event}) \times \text{ Monetary value of Event}$$

where the sum is over all possible events. The EMV of a project can be used as a decision criterion for choosing between different projects and has applications in a large number of situations. This is illustrated by the following example.

A small company is trying to decide how to launch a new and innovative product. It could go for a direct approach, launching onto the whole of the domestic market through traditional distribution channels, or it could launch only on the internet. A third option exists where the product is licensed to a larger company through the payment of a licence fee irrespective of the success of the product. How should the company launch the product? The company has done some initial market research and the managing director believes the probability of the product being successful can be classed into three categories: high, medium or low. She thinks that these categories will occur with probabilities 0.2, 0.35 and 0.45 respectively and her thoughts on the likely profits (in £K) to be earned in each plan are

	High	Medium	Low
Direct	100	55	-25
Internet	46	25	15
Licence	20	20	20

The EMV of each plan can be calculated as follows:

$$\begin{aligned} EMV(\text{Direct}) &= 0.2 \times 100 + 0.35 \times 55 + 0.45 \times (-25) = \pounds 28\text{K} \\ EMV(\text{Internet}) &= 0.2 \times 46 + 0.35 \times 25 + 0.45 \times 15 = \pounds 24.7\text{K} \\ EMV(\text{Licence}) &= 0.2 \times 20 + 0.35 \times 20 + 0.45 \times 20 = \pounds 20\text{K}. \end{aligned}$$

On the basis of expected monetary value, the best choice is the Direct approach.

In this example we have to make a decision. When we include a decision in a probability tree we use a rectangular node, called a *decision node* to represent the decision. The diagram is then called a *decision tree*. There are no probabilities at a decision node but we evaluate the expected monetary values of the options. In a decision tree the first node is always a decision node. There may also be other decision nodes. If there is another decision node then we evaluate the options there and choose the best and the expected value of this option becomes the expected value of the branch leading to the decision node.

6.5 Exercises 6

1. A company has installed a new computer system and some employees are having difficulty logging on to the system. They have been given training and the problems which arose during training were recorded and their probabilities calculated as follows:

   - An employee has a probability of 0.9 of logging on successfully on the first attempt.
   - If the employee logs in successfully then the employee will also be successful on each later attempt with probability 0.9.
   - If the employee tries to log in and is not successful then the employee loses confidence and the probability of a successful log-in on later occasions drops to 0.5.

   Use a tree diagram to find the following probabilities:

   (a) An employee successfully logs on in each of the first three attempts.

   (b) An employee fails in the first attempt but is successful in the next two attempts.

   (c) An employee logs on successfully only once in three attempts.

   (d) An employee does not manage to log on successfully in three attempts.

2. The owner of a small business has the right to have a retail stall at a large festival to be held during the summer. She judges that this would either be a success or a failure and that the probability that it is a success is 0.4. If the stall was a success, the net income from it would be £90,000. If the stall was a failure, there would be a net loss of £30,000 from it. To help to make the decision, the owner could pay for market research. This would cost £5,000. The market research will either give a positive indication or a negative indication. The conditional probability that it gives a positive indication, given that the stall will actually be a success, is 0.75. The conditional probability that it gives a positive indication, given that the stall will actually be a failure, is $1/3$.

   The owner has various options:

   - Do nothing.
   - Go ahead without market research.
   - Pay for the market research.
   - Sell her right to a stall for £10,000.

   If she pays for the market reserach then, depending on the outcome, she can:

   - Do nothing more.
   - Go ahead.
   - Sell her right to a stall. If the market research gave a positive indication the price would be £35,000. If the market research gave a negative indication the price would be only £3,000.

   What should she do?

   This is a fairly complicated question so it is best to tackle it in stages.

(a) using a probability tree for the success or failure of the stall and the market research outcome, find the following probabilities.

  i. The probability of a positive market research outcome.
  ii. The probability of a negative market research outcome.
  iii. The conditional probability of a successful stall given a positive market research outcome.
  iv. The conditional probability of a failure given a positive market research outcome.
  v. The conditional probability of a successful stall given a negative market research outcome.
  vi. The conditional probability of a failure given a negative market research outcome.

(b) Represent the owner's decision problem using a tree diagram.

(c) Suppose that the owner has the market research done and that the outcome is positive. Evaluate the expected monetary values, under these circumstances, of the three options:

  - Sell.
  - Go ahead.
  - Do nothing more.

  and hence find what she should do if these circumstances arise.

(d) Suppose that the owner has the market research done and that the outcome is negative. Evaluate the expected monetary values, under these circumstances, of the three options:

  - Sell.
  - Go ahead.
  - Do nothing more.

  and hence find what she should do if these circumstances arise.

(e) Hence find the expected monetary value of the initial option of "Pay for the market research."

(f) Find the expected monetary values of the three other inital options:

  - Do nothing.
  - Go ahead without market research.
  - Sell her right to a stall for £10,000.

(g) Determine the owner's best strategy.