2. Basic Probability

In this section:

- What is a Probability?
- The 3 Laws of Probability
- Conditional Probability
- Independence

Some definitions:

An experiment is a random process or object

An outcome is the result of the experiment

An event is an outcome or a set of outcomes.

The set of all possible outcomes is called the *Sample Space*.

For a throw of a standard Die the set is: {1,2,3,4,5,6}.

For the status of an email message it might be:

spam, not spam

For the result of a men's tennis match the sample space might be:

{ 3:0, 3:1, 3:2, 2:3, 1:3, 0:3} - interest in sets only, best of five sets.

Sample spaces must contain all possible events but can contain “extra events” that do not occur in practice:

Ex.

Heights of people in cm.

A convenient sample space here is the positive integers – even though we are unlikely to come across people 500cm tall!

An even more convenient one is the real line.

Result of a football match A:B where A and B are non-negative integers.

The definition of the sample space may depend on our ability to assign probabilities to the *elementary events* (individual points) in it.

Two dice:

Often only the sum is of interest:

Possibilities 2,3,..12.

Outcomes on 2 dice:

(1,1) (1,2)… (1,6)

(2,2) (2,3)…(2,6)

…

… (6,6).

21 possibilities (Check it!)

However…
Two distinct dice: first, second or blue, red treat (1,2) and (2,1) as different outcomes a total of 36 outcomes.

The last one is finer than the ones above. It is the one that we can assign probabilities to the events. Equally likely .

We can compute the probabilities for the other outcomes by the rules of probability.

2.1 What is a Probability?

- Probability is a number describing the 'chance' or 'likelihood' of something happening.

  i.e. throw a coin. P(heads) = 0.5

Probability is an axiomatic system – like geometry. We can define rules for manipulating it without specifying what it exactly is.

Any measure of likeliness that satisfies the following will do as a probability:

1. $0 \leq P(A) \leq 1$ - probabilities shall be between 0 and 1.

2. P(certain event) = 1 <=> P(impossible event)= 0

3. If A and B cannot occur together – *mutually exclusive .*

$$P(A \text{ or } B) = P(A) + P(B)$$

In this definition we are not told how to calculate probability.

There are three methods of doing this.

1. Equally likely events – in this method we argue from symmetry that certain events have the same chance of occurring.

Ex. ideal (perfect) coin
two outcomes {H,T}

P(H)=P(T) by symmetry

P(H or T) =1 as the certain event.

P(H or T) = P(H) + P(T)

Thus P(H) = P(T) =0.5.

We can only use this method if we can identify equally likely events.

Man Utd play Sheffield Utd in the Cup
Three outcomes Win, Draw, Loss (for Man U)
But it would not be sensible to suggests the the 3 outcomes have the same probability.

Generally

$$P(\text{event}) = \frac{\text{number of outcomes where event happens}}{\text{total number of outcomes}}$$

Coin: outcomes where heads happens = 1
total number of outcomes = 2

=> P(heads) = 1/2

6-sided Die:

6 equally likely outcomes P(each outcome) =1/6
Events can be compound:
P(even number) = 3/6 =1/2

3 because three elementary outcomes “2”, “4” and “6” correspond to the event “even” occurring.

The '*Frequency*' model

P(event) = proportion of times event has been observed to occur in long run.

Coin: we've thrown a coin many times in the past and 'heads' has occurred half of the time

⇨ P(heads) = ½

But also we can use this method to estimate events where we have no obvious symmetry:

Exams: Over the past ten years, 350 students out of 7000 have failed an exam

=> P(student fails) = 350/7000 = 0.05

Still can’t calculate P(Man Utd win) directly.

But in fact the method can be used here –

A complex mathematical model for team ability can be constructed. Parameters of this model are estimated using past data (results against other teams). – this implicitly uses the frequency model.

Generally:

Conduct n identical experiments, event E occurs m times in n attempts

$$P_n(E) = \frac{m}{n}$$

The probability of E is defined:

$$P(E) = \lim_{n \to \infty}(P_n(E))$$

If n is large we hope that the value of $P_n(E) = \frac{m}{n}$ estimates (closely) P(E).

Subjective probability

P(event) = your personal *belief* in the chance of the event happening

Use your own knowledge, experience, to assign P

Coin: I believe the coin to be fair, so P(heads) = $\frac{1}{2}$.

This method is gaining in popularity. It can be shown (mathematically) that consistent, unique probabilities values do exist – there is however a practical problem of deciding what they are.

Whatever method is used for calculating probabilities we need methods for manipulating them.

Combining events

Remarks.
In maths “or” gets replaced by the union: U
“and” by the intersection: .∩

Mutually exclusive events correspond to disjoint sets of the sample space.

P(A and B) = P(impossible) =0

For general events we have:

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

P(A or B) =P(A) + P(B) – P(A and B)

Example:
A university has 10000 students. 6000 of whom are male. 400 students are in Computer Science, 250 of whom are male.

A student is chosen at random what are the probabilities of:

P(male) =? P(female) =?

P(In CS) =?

P(female or in CS) =?

P(female and in CS)=?

P(male and in CS and will fail Stats exam) =?

Solution

Chosen at random means that each student has an equal chance of being chosen, the chosen student is equally likely to be any of the 10000 – so equally likely to defn. applies by default.

P(male) = 6000/10000 =0.6
P(female) = P(not(male)) = 1 – 0.6 =0.4

P(not(A)) = 1 – P(A).

P(In CS) = 400/10000 = 0.04

P(female or in CS) = P(female) + P(in CS) – P(female and in CS).

P(female and in CS) = #females in CS/10000
= (400-250)/10000
= 0.015

so
P(female or in CS) = 0.4 + 0.04 –0.015=0.435

Insufficient data to compute this as we do not even know how many CS students take a stats exam never mind their chances of passing. But we can say for sure that:

$0 \le P(...) \le 0.025$ = P(male and in CS)

adding an “and” condition can only reduce the probability.
Adding an “or” condition can only increase it.

P (A and B) depends on the relationship between these events. Does the occurrence of A raise, lower or leave the same the chances of B occurring.

To study this we use :

2.2 Conditional Probability

- P(A) means "Probability of A happening"
- P(A | B) means "Probability of A happening *given* that B has already occurred"

  This is called a *conditional* probability: prob. of A conditional on B. or just “A given B”

Definition:

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

This implies:

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

Example – Students again.

$$P(CS \mid female) = \frac{P(CS \cap female)}{P(female)} = \frac{0.015}{0.4} = 0.0375$$

$$= \frac{\#\, females \text{ in } CS}{\#\, females} = \frac{150}{4000}$$

The possibilities (sample space) is reduced to females only.

$$P(female \mid CS) = \frac{P(CS \cap female)}{P(CS)} = \frac{0.015}{0.04} = 0.375$$

Note that the first probability is less than P(CS) and the second less than P(female) . Being female discourages computer science and vice versa. – The proportion of females in CS is smaller than the proportion of females in the whole university.

A special case is when there is no effect

2.3 Independence

Take two events A and B

If the occurrence of one does not affect probability of the other, they are called *Independent*

A, B independent <=> $P(A \mid B) = P(A)$,

and

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

Or equivalently:

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

Studying dependence, through conditional probability, gives us an insight relationships between events. However the analysis becomes very complicated when more than two events are involved.

Analysing dependent data gets very difficult. Thus data collection methods are usually devised so that the observations are independent.

Or that the dependence has a very simple structure that is clearly understood.

Some examples.

1.Throwing a dice twice

As long as the second throw is reasonable – ie no effort is made to deliberately use the position on the first throw to affect the second these should be independent.
Then the probabilities of all events concerning the two dice are easily computed.

P(two sixes) = P(first dice=6)*P(second dice =6)
= 1/6 * 1/6
= 1/36

2. Students
We have seen that the events CS and female are not independent because of the exact numbers of CS and females in the university.

If the proportion of females in CS was exactly the same as in the whole university ie. If there were 160 females in CS (both proportions=40%).
Then these events would be independent.

Learning that the randomly chosen student is female would not carry any info regarding whether she was CS or not.

On the other hand if all CS were male then $P(CS \cap female) = 0$ and the effect would be quite dramatic.

3. Velocity of light data.

Although we cannot be sure about this, any analysis of the data would assume that the measurements are independent of one another.

(Events here are: measurement = particular value)

The assumption is that the kth measurement was NOT adjusted by looking at the value of the previous ones – it would be poor scientific practice to do so.

Checks could be carried to see if there is any evidence of data manipulation.

4.Bugs is software

Here the data by their very nature are dependent – if a large number of bugs are discovered early there are fewer to be discovered later.

Any analysis has to take account of this – however the dependence here is well defined and relatively simple.

5.PCs

In the analysis of this experiment it would be assumed that each PC survives or fails independently of the others. The experimenter must ensure that this is true. If it is not, no clear way to come up with any conclusions.

How might independence be violated in this example?

In practice few data satisfy the independence conditions. However the dependence is so slight that the analyses conducted are approximately correct.

A nice example are pseudo random number generators –

RNGs generate sequences of numbers according to an algorithm. Given a seed the sequence is predetermined. Thus r_n determines r_{n+1} – the values of r are clearly dependent. However the dependence structure is sufficiently “strange” and complicated that for practical purposes the sequence of numbers may be regarded as independent!

Random numbers are used for simulating physical processes. The dependence in RNGs is so unlike any physical process that it doesn’t bias the simulation.

Statistical tests of randomness look for physical type dependences and departures and (good) RNGs pass such tests.