MS4S08 – Applied Statistics for Data Science

Introduction to Statistics

Dr Penny Holborn
penny.holborn@southwales.ac.uk

Probability

We use the term probability in everyday speech.

“What is the probability that it will rain on Saturday?”

“What is the probability that Wales will win the next Rugby World Cup?”

“What is the probability that I will win the lottery jackpot on Saturday?”

“If a fair coin is tossed what is the probability that it falls ‘Heads’?”

In the last example most people would accept that the answer is 0.5, but the answers to the other questions are more complicated.

University of South Wales
Prifysgol De Cymru

Probability

The mathematical definition of probability agrees with the everyday usage, but also allows us to find the probability of more complicated events by using the rules governing the addition, subtraction, multiplication and division of probabilities of simple events.

For example, we can find the probability that 10 tosses of a fair coin give 7 'Heads'.

We call the process of observation or measurement an **Experiment**.

Example

a. Noting whether a coin falls 'Heads' or 'Tails' when tossed;
b. Recording the value shown when a die is rolled;
c. Recording the results of a laboratory experiment;
d. Recording the responses to a questionnaire.

Probability

Any result of the experiment is called an **Outcome.**

Example

a. Coin falls 'Heads';
b. Die shows 5;
c. Result is 15.67 %;
d. Answer to a question is ‘Agree’.

Any outcome or set of outcomes is called an **Event**.

Example

a. {Result of coin toss is 'Heads'};
b. {value shown on die is even};
c. {result falls between 10.00 % and 19.99 %};
d. {Response is ‘Agree’ or ‘Strongly agree’}.

University of South Wales
Prifysgol De Cymru

Probability

A set of events is said to be **mutually exclusive** if they are such that no two of them can occur together as the result of the same experiment.

A set of events is **exhaustive** if they include all the possible outcomes of an experiment.

Example

a. Coin toss- **the events {Head} and {Tail} are clearly exclusive and exhaustive.**

b. Rolling a die - **the events {3 or more} and {5 or less} are exhaustive but not exclusive.**
  - **the events {3 or less} and {5 or more} are exclusive but not exhaustive.**

The definition of probability can be developed in 2 ways.

University of South Wales
Prifysgol De Cymru

Definition of Probability (a)

Suppose there are *N* possible outcomes of an experiment all equally likely to occur, *n* of which satisfy the conditions of an event E.

Then the probability that E will occur is defined as $\mathrm{P(E)} = \frac{n}{N}$

Example
One letter is selected at random from the English alphabet.
What is the probability that it is a vowel?

N = number of letter in the English alphabet = 26
n = number of vowels = 5

P(vowel) = ${}^{5}/_{26}$

Definition of Probability (a)

The following data gives the hair and eye colour of a group of 200 people.

		HAIR		
		Dark	Medium	Fair
EYE	Blue	2	40	28
	Brown	45	40	15
	Other	3	20	7

a) Find the probability that a person selected at random from this group will have blue eyes.
b) Find the probability that a person selected at random from this group will not have dark hair.

Definition of Probability (a)

		HAIR			
		Dark	Medium	Fair	Total
EYE	Blue	2	40	28	70
	Brown	45	40	15	100
	Other	3	20	7	30
	Total	50	100	50	200

a) *N = number of people= 200*
n = number with blue eyes= 70

P(blue eyes) = $^{70}/_{200} = 0.35$

University of South Wales
Prifysgol De Cymru

Definition of Probability (a)

		HAIR			
		Dark	Medium	Fair	Total
EYE	Blue	2	40	28	70
	Brown	45	40	15	100
	Other	3	20	7	30
	Total	50	100	50	200

a) *N = number of people= 200*
n = number without dark hair= 150

P(without dark hair) = ${}^{150}/_{200} = 0.75$

Definition of Probability (b)

Suppose that in a series of N identical trials an event E occurs n times.

As N becomes large the proportion of occurrences $\frac{n}{N}$ will be very close to P(E).

So we can define P(E) = $\frac{n}{N}$.

We see that, for any event E, $0 \leq P(E) \leq 1$

with P(E) = 0 if and only if E cannot occur (impossible event)

and P(E) = 1 if and only if E must occur (certain event)

Definition of Probability (b)

Independent Events

Two events A and B are said to be independent if the probability that B occurs is not affected by whether or not A occurs (and vice versa).

Example

Two cards are drawn at random from a standard pack.

Let A represent the event {first card is an Ace} and let B represent the event {second card is Ace}.

We assume the first card is not replaced before the second draw.

Therefore these are not independent events.

Definition of Probability (b)

If A & B are two **independent** events then the probability that they both occur is given by:

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

This can also provide a test for independent events

Two events A and B are said to be independent if and only if

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

Definition of Probability (b)

Let A denote the event that a person selected at random from the group has dark hair.

P(A) =$^{50}/_{200}$

Let B denote the event that a person selected at random from the group has brown eyes.

P(B) =$^{100}/_{200}$

The event {A and B} corresponds to a person having both dark hair and brown eyes, thus

P(A and B) = $^{50}/_{200} * ^{100}/_{200}$ = 0.125

But from the table we can see that P(A and B) = $^{45}/_{200}$ = 0.225.

Hence the events A and B cannot be said to be independent.

Definition of Probability (b)

The concept of independence may be extended to more than 2 events.

Three events A, B and C are said to be independent if the probability that A occurs is not affected by whether or not B and/or C occur (and similarly for B and C).

If A, B & C are independent events then the probability that they all occur is given by:

P(A and B and C) = P(A).P(B).P(C)

Definition of Probability (b)

For any two events, the probability that one (or both) of A or B occur is given by:

P(A or B) = P(A) + P(B) – P(A and B)

If A & B are mutually exclusive events then P(A and B) = 0 so we have

P(A or B) = P(A) + P(B)

We can also deduce that:

P(A does not occur) = 1 - P(A)

Definition of Probability (b)

Example

Out of 200 students taking an introductory statistics course, 140 passed the examination, 160 passed the coursework and 124 students passed both the exam and coursework.

The lecturer (who is a kind **woman**) decides that a pass will be awarded to any student who has passed either the examination or the coursework (or both).

What is the probability that a randomly selected student has passed the course?

Let event A = passed the exam

Let event B = passed the coursework

P(A) = 140/200

P(B) = 160/200

P(A and B) = 124/200

P(A or B) = P(A) + P(B) – P(A and B)

= 0.88

University of South Wales
Prifysgol De Cymru

Definition of Probability (b)

Example

Two people, A and B each fire a single shot at a target. A hits a target, on average, 80 times out of 100 shots whilst B does so 90 times out of 100 shots, on average.
Find the probability that the target is hit.

P(A) = 80/100 = 0.8
P(B) = 90/100 = 0.9

P(A hits and B hits)= 0.8 x 0.9 = 0.72
P(A hits and B misses) = 0.8 x 0.1 = 0.08
P(A misses and B hits) = 0.2 x 0.9 = 0.18

Therefore P(hit) = 0.72 + 0.08 + 0.18 = 0.98

Or P(hit) = 1 – P(A misses and B misses) = 1 – (0.2 x 0.1) = 1 – 0.2 = 0.98

University of South Wales
Prifysgol De Cymru

Conditional Probability

When events A & B are **not independent** we write the probability of B occurring given that A has already occurred as $P(B|A)$.

The rule for calculating P(A and B) now becomes:

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

Hence $\quad P(B|A) = P(A \text{ and } B) / P(A) \quad$ and $\quad P(A|B) = P(A \text{ and } B) / P(B)$

If A and B are independent then we see that:

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

So that we again have $P(A \text{ and } B) = P(A).P(B)$.

Conditional Probability

Back to previous example

		HAIR			
		Dark	Medium	Fair	Total
EYE	Blue	2	40	28	70
	Brown	45	40	15	100
	Other	3	20	7	30
	Total	50	100	50	200

Suppose it is known that a person has dark hair.
What is the probability that they also have brown eyes?

P(brown eyes/dark hair) = P(brown eyes and dark hair)/P(dark hair)
= (45/200)/(50/200) = 9/10 = 0.9

Exercises 7

It is time to work through some exercises yourselves.

Please feel free to ask any questions.

University of South Wales
Prifysgol De Cymru

Exercises 7

1. A letter is chosen at random from the word "MATHEMATICAL".
Find the probability that it is:-

(i) a vowel ; (ii) an M or a C ; (iii) not a T .

String or substring	Number of characters
MATHEMATICAL	12
AEAIA	5
MMC	3
TT	2

(i) P(vowel) = 5/12 = 0.416667
(ii) P(an M or a C) = 3/12 = 0.25
(iii) P(not a T) = 1 – 2/12 = 10/12 = 5/6 = 0.83333

Exercises 7

2. An estate agent classifies his houses for sale by area (A, B, C) and also by the number of bedrooms. The following pattern is found:

	Area A	Area B	Area C
2 Bedrooms	10	10	10
3 Bedrooms	20	50	60
4+ Bedrooms	20	10	10

(a) If a client chooses a house at random, find the probability that:-
(i) it has 3 bedrooms ; (ii) it is in area B or C .

(b) Are the events {House has 3 bedrooms} and {House is in Area B} independent?

Exercises 7

	Area A	Area B	Area C	Total
2 Bedrooms	10	10	10	**30**
3 Bedrooms	20	50	60	**130**
4+ Bedrooms	20	10	10	**40**
Total	**50**	**70**	**80**	**200**

a)

(i) P(3 bedrooms) = 130/200 = 13/20 = 0.65

(ii) P(in Area B or C) = (70+80)/200 = 150/200 = 3/4 = 0.75

Exercises 7

If the events {3 bedrooms} and {Area B} *were* independent then we would expect that:

P({3 bedrooms} AND {Area B}) = P(3 bedrooms) × P(Area B).

We note that:
P(3 bedrooms) = 130/200 = 13/20 = 0.65 (as above)
P(Area B) = 70/200 = 0.35

Therefore P(3 bedrooms) x P(Area B) = 0.65 × 0.35 = 0.2275

However, we see from the table that P({3 bedrooms} AND {Area B}) = 50 / 200 = 0.25.

The probabilities are different indicating that the events are **not independent**.

University of South Wales
Prifysgol De Cymru

Exercises 7

3. A firm submits tenders for four separate contracts. The probabilities of acceptance are 0.3 , 0.6 , 0.5 and 0.4 respectively. If the chances of success are independent, find the probability that:-

(i) all the tenders are accepted ;
(ii) none of the tenders are accepted ;
(iii) at least one of the tenders is accepted.

(i) P(all accepted) = $0.3 \times 0.6 \times 0.5 \times 0.4 = 0.036$

(ii) P(none accepted) = $(1 - 0.3) \times (1 - 0.6) \times (1 - 0.5) \times (1 - 0.4) = 0.056$

(iii) P(at least one accepted) = 1 – P(none accepted) = 0.944

University of South Wales
Prifysgol De Cymru

Exercises 7

4. Two marksmen whose probabilities of hitting a target are 0.7 and 0.8, respectively each fire one shot. What is the probability that:-

(a) both hit the target;
(b) only the first hits the target ;
(c) at least one hit the target ?

Independent events.

(a) P(Both hit) = $0.7 \times 0.8 = 0.56$

(b) P(only first hits) = $0.7 \times 0.2 = 0.14$

(c) P(at least one hit) = $0.7 \text{ x } 0.8 + 0.7 \times 0.2 + 0.3 \times 0.8 = 0.56 + 0.14 + 0.24 = 0.94$

Exercises 7

5. A large batch of components contains 10% which are defective. If the components are tested one at a time, find the probability that :-

(i) the first defective item will be the third to be tested;
(ii) when three items are tested exactly one defective will be found.

(i) P(first defective will be third tested) = $0.9 \times 0.9 \times 0.1 = 0.081$

(ii) P(one defective from 3) = $(0.9 \times 0.9 \times 0.1) + (0.9 \times 0.1 \times 0.9) + (0.1 \times 0.9 \times 0.9) = 0.243$

Exercises 7

6. Four letters are written and four envelopes addressed. If the letters are allocated randomly to the envelopes, find the probability that:-

(i) all are in the correct envelopes ;
(ii) exactly one is in the correct envelope.

(i) $P = 1/4 \times 1/3 \times 1/2 \times 1 = 1 / 24 = 0.041667$

(ii) P(1 letter placed in the correct envelope and rest incorrect) =
P(letter 1 placed in envelope 1 and rest incorrect) = $1/4 \times 2/3 \times 1/2 \times 1 = 1/12$
+ P(letter 2 placed in envelope 2 and rest incorrect) = $1/4 \times 2/3 \times 1/2 \times 1 = 1/12$
+ P(letter 3 placed in envelope 3 and rest incorrect) = $1/4 \times 2/3 \times 1/2 \times 1 = 1/12$
+ P(letter 4 placed in envelope 4 and rest incorrect) = $1/4 \times 2/3 \times 1/2 \times 1 = 1/12$

Therefore P(1 letter placed in the correct envelope and rest incorrect) = $4 \times 1/ 12 = 1 / 3$