4 Fundamentals of Probability

We all have an intuitive idea of the meaning of the “chance”, “probability”, or “likelihood” of an event occurring.

- What’s the probability of rolling a 6 on one die?
- You toss a coin 100 times. What’s the probability of a run of 5 or more heads occuring?
- What’s the probability of observing `TATAAA` in a sequence of 100 base pairs of DNA?
- What’s the probability that an internet user will follow a link from an advert?
- What’s the probability that I’ll live until I’m 65?

The values we assign these examples and our interpretation of them questions what we really mean by “probability”. In this section we go from our intuitive notions of probability and examples like those above to define a formal mathematical notion of probability. With this in place, we will be in a position to calculate probabilities and thereby quantify our uncertainty about events or hypotheses of interest, e.g. “how strong is the evidence?”

4.1 Events

4.1.1 Definitions

We start to formalize our notion of probability by considering an *experiment*, that is, a process that generates data. Although we generally cannot predict the outcome of such an experiment, we assume we can list the collection of possible outcomes. The set of all possible outcomes is called the *sample space* and is denoted S. The size of the sample space can be either finite or infinite.

Define sample spaces for the following experiments:

- Tossing a coin one time -
- The number of packets of cereal purchased by a customer in 13 weeks -
- The survival time of leukaemia patients -
- The number of viable cells out of 100 cell cultures -
- Taking the age of a random individual (in the world) -

Example:
A die is thrown once. There are six possible outcomes and $S = \{1, 2, 3, 4, 5, 6\}$. We might be interested in the following events:

1. the outcome is the number 1;
2. the outcome is an even number;
3. the outcome is the number 6 or an odd number;
4. the outcome is an even number and is less than 4;
5. the outcome is not even.

Each of these events corresponds to a *subset* of S:

1. $A = \{1\}$,
2. $A = \{2, 4, 6\}$,
3. $A = \{6\} \cup \{1, 3, 5\} = \{1, 3, 5, 6\}$
4. $A = \{2, 4, 6\} \cap \{1, 2, 3\} = \{2\}$
5. $A = \{2, 4, 6\}^c = \{1, 3, 5\}$.

From now on we will think of *events* as being *subsets* of the sample space S. Remember the distinction between events and outcomes: outcomes are elements of S; events are subsets and can thus represent several different outcomes. Event A *occurs* if the outcome of the experiment, s, is contained in the event A (i.e. $s \in A$, where $\in$ is read "is an element of").

Exercise:
Write down the following events. ✐

- For the cereal purchases experiment $S = \{0, 1, 2, 3, \ldots\}$, specify the events:
  - A_1 : customer buys at most 2 packets -
  - A_2 : customer buys at least 1 packet -

4.1.2 Set operations

The examples above showed how different events can be described using *set operations* such as the union, intersection, and complement.

The *union* of two events A, B is the event consisting of all the outcomes that are either in A or in B or in both. It is denoted $A \cup B$.

The *intersection* of two events A, B is the event consisting of all the outcomes that are in both A and B. It is denoted $A \cap B$.

The *complement* of an event A (with respect to the sample space S) is the event consisting of all the outcomes in S that are not contained in A. It is denoted A^c or sometimes $\bar{A}$.

You might find the following table of notation useful.

Notation	Definition	Meaning
S	Sample space	Set of all outcomes – a certain event
$A \subset S$	An event	A sub-collection of possible outcomes
A^c	Complement of A	Event that an outcome in S but not in A occurs
$A \cap B$	Intersection of A and B	Both A and B occur
$A \cup B$	Union of A and B	Either A or B or both occur
$\emptyset$	Empty set	Impossible event

Note that in this table the symbol $\subset$ reads "is a subset of". In general we can write $A \subset B$ if every outcome in A is also in B.

4.1.3 Disjoint events

Remember, we say that event A *occurs* if the outcome of the experiment, s, is contained in A. Two events A and B are *disjoint* or *mutually exclusive* if they cannot both occur, so their intersection is empty:

$$A \cap B = \emptyset.$$

For any event A, the events A and A^c are disjoint, and their union is the whole of the sample space:

$$A \cap A^c = \emptyset \quad \text{and} \quad A \cup A^c = S.$$

Event A *implies* event B, denoted $A \Rightarrow B$, if B occurs whenever A occurs i.e. $A \subset B$.

4.2 Axioms of probability

Given an experiment and a sample space S, we want to be able to assign to each event A a number, $\Pr(A)$, called the *probability of* A, which quantifies the chance of event A occuring. In order to ensure that our probability assignments are sensible and agree with our intuitive notions of probability, they should satisfy the following basic properties called the *axioms of probability*:

1. $\Pr(A) \geq 0$ for any event A;
2. $\Pr(S) = 1$;
3. If $A_1, A_2, A_3, \dots$ is an infinite collection of disjoint events then

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \dots) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \dots$$

Axiom 1 reflects the intuitive idea that probabilities of events should not be negative. The intuition behind axiom 2 is that the sample space S contains all possible outcomes of the experiment and so S is an event which must occur when an experiment is performed. Axiom

3 ensures that if we are working with disjoint events, the probability that at least one of them occurs can be obtained by adding the individual probabilities. It is possible to derive many useful results from these axioms, for example, one can verify the intuitive results that $\Pr(\emptyset) = 0$ and $\Pr(A) \leq 1$ for any event A.

4.3 Interpretations of probability

The axioms of probability do not completely determine an assignment of probabilities to events. There are two main ways of interpreting probabilities, both of which are consistent with the axioms above. The traditional viewpoint is the "frequentist" or "limiting relative frequency" interpretation. However, over the past few decades, the "subjective" viewpoint, which is used by Bayesian statisticians, has become increasingly popular. In this introductory course we will focus on the former interpretation but you will learn more about the "subjective" viewpoint in later courses.

4.3.1 Probability as a limiting relative frequency

Suppose that we repeat an experiment independently a large number of times, and record which events occur. Some events occur more than others, but as the experiments go on, the *proportion* of times an event A occurs converges to a fixed value. This value is the *probability* of the event A:

$$\Pr(A) = \frac{\text{Number of times } A \text{ occurs}}{N}$$

as the number of trials N increases. Based on this definition, we can see, for example, that $\Pr(A)$ must be non-negative (axiom 1) and that $\Pr(S) = 1$ (axiom 2).

4.4 Equally likely outcomes

Suppose that S is finite with n elements. If it is reasonable to assume that every outcome in S is **equally likely** then

$$\Pr(\{s\}) = \frac{1}{n}$$

for any outcome s. The probability of an event A is then determined by the number of elements it contains:

$$\Pr(A) = \frac{\text{Number of elements in } A}{n}$$

Example:
Suppose that a fair coin is thrown twice, and the results recorded. The sample space is

$$S = \{HH, HT, TH, TT\}$$

Define A to be the event *head on the first toss*, and B to be the event *head on the second toss*

Find

$\Pr(A)$	
$\Pr(B)$	
$\Pr(A \cup B)$	
$\Pr(A \cap B)$	

Note: it may be easier to think about $\Pr(A \cup B)$ as $\Pr(A \text{ or } B)$ and $\Pr(A \cap B)$ as $\Pr(A \text{ and } B)$.

Exercise:
Suppose that we throw two dice. The sample space is

$$S = \{(1,1),(1,2),(1,3),\dots,(5,6),(6,6)\}$$

in total 36 different events. Define A to be the event *Number one on the first roll*, and B to be the event *Total is 5*. Fill in another table like the example above.

Find

$\Pr(A)$	
$\Pr(B)$	
$\Pr(A \cup B)$	
$\Pr(A \cap B)$	

4.5 Specifying probabilities numerically

We have seen that the probability of an event is a number between 0 and 1 which can be interpreted as the ratio of the "number of successes" to the "number of trials". Probabilities can also be specified via *odds*:

- Odds on = (number success) : (number fail)
- Odds against = (number fail) : (number success)

Odds are frequently used by bookmakers to represent the value of various bets. The above system gives what are known as *fractional* odds (as opposed to the American or decimal systems), and as a general rule the "odds against" are reported (i.e. a bet being advertised at 2/1 would return 3x the amount the customer were to stake should the bet be successful).

Example:
No course in probability is complete without an example involving picking coloured counters out of a bag. Suppose in this case we have 9 counters in a bag, 3 red and 6 blue, and are interested in the odds that we select a red counter (denoted R).

Counters in bag	RRRBBBBBB
Odds on R	
Odds against R	
Probability of R	

4.6 Adding probabilities

We return to our pool of probability cliches and consider the case where we roll a die and are interested in several outcomes. First we wish to calculate the probability the roll produces either a 1 or a number greater than 3. From probability axiom 3 we know we can work out the probability of any of a series of disjoint events occurring by simply adding the probabilities of the individual events,

$$\begin{aligned}\Pr(\text{Roll a } 1 \cup \text{Roll greater than } 3) &= \Pr(\text{Roll a } 1) + \Pr(\text{Roll greater than } 3)\\ &= \frac{1}{6} + \frac{3}{6}\\ &= \frac{4}{6} \quad \left(= \frac{2}{3}\right).\end{aligned}$$

This all seems very sensible. Suppose now we are interested in a pair of events which are not disjoint, say if we are interested in the probability that we roll an odd number or a number greater than 3. We can try and attempt the formula from axiom 3 again,

$$\begin{aligned}\Pr(\text{Roll odd or Roll greater than } 3) &= \Pr(\text{Roll odd}) + \Pr(\text{Roll greater than } 3)\\ &= \frac{3}{6} + \frac{3}{6}\\ &= \frac{6}{6} \quad (= 1).\end{aligned}$$

This would seem to suggest it is certain that our roll will be either odd or be over 3, and that these two events make up the entire sample space S. However we know this is not true, rolling a 2 is possible, and would not satisfy either of the two outcomes we are investigating. So where did we go wrong?

We consider the outcomes which make up the two events, "roll an odd number" and "roll greater than a 3",

$$A = \text{Roll odd} = \{1, 3, 5\} \qquad B = \text{Roll greater than } 3 = \{4, 5, 6\}.$$

Immediately we see the problem, by simply adding the two probabilities we have *double counted* the outcomes which satisfy both events, in this example that we roll a 5. This is not a problem for disjoint events as by definition there is no outcome which satisfies both events, however for non-disjoint events we must remember to subtract the probability that an outcome occurs which satisfies both events, in order to avoid this double counting. This gives rise to the *addition rule* given below,

$$\boxed{\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)}$$

The second axiom of probability states that:

$$\Pr(S) = 1.$$

By the definition of a complement we get:

$$\Pr(A) + \Pr(A^c) = \Pr(S) = 1$$

and so

$$\boxed{\Pr(A^c) = 1 - \Pr(A)}$$

Exercise:
Given $\Pr(A) = 0.3$, $\Pr(A \cup B) = 0.45$, and $\Pr(B^c) = 0.75$, work out:

1. $\Pr(A \cap B)$, and
2. $\Pr(A^c \cup B)$

Solution.

4.7 Conditional probability and the Multiplication Rule

We return to our example of choosing coloured counters out of a bag, and recall we had 9 counters, 3 red and 6 blue, and were interested in the probability of selecting a red counter. Suppose now rather than simply be curious as to the probability of selecting a red counter followed by a blue. If we were to return the first counter selected to the bag our events would be independent (the colour of the first colour we select doesn't affect the second selection at all) and so we can do this simply by multiplying the probability of selecting a red (R) by the probability of selecting a blue (B),

$$\begin{aligned}\Pr(R \cap B) &= \Pr(R) \times \Pr(B)\\ &= \Pr(R) \times \Pr(B)\\ &= \frac{3}{9} \times \frac{6}{9}\\ &= \frac{18}{81} \quad \left(= \frac{2}{9}\right).\end{aligned}$$

Now suppose we don't replace the first counter we select, our selections would no longer be independent, since if for instance the first counter we selected was a red, there would be 8 counters remaining in the bag, with 6 of them blue, making the probability of selecting a blue counter $\frac{6}{8}$. This concept of the probability of an event given another event has already occurred is known as *conditional probability* and is denoted $\Pr(A|B)$, which means "the probability that event A occurs given event B has already occurred". This gives rise to the *multiplication rule*,

$$\boxed{\Pr(A \cap B) = \Pr(A|B) \times \Pr(B) \quad \text{for } \Pr(B) > 0}.$$

Hence we can return to our counter choosing game, where we wish to know the probability of selecting a red counter and then a blue counter without replacing the first counter. Using the multiplication rule we get,

$$\begin{aligned}\Pr(R \,\&\, B) &= \Pr(R) \times \Pr(B|R)\\ &= \Pr(R) \times \Pr(B)\\ &= \frac{3}{9} \times \frac{6}{8}\\ &= \frac{18}{72} \quad \left(= \frac{1}{4}\right).\end{aligned}$$

Rearranging the multiplication rule by dividing both sides by $\Pr(B)$ gives the formula for conditional probability,

$$\boxed{\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{for } \Pr(B) > 0}.$$

The requirement $\Pr(B) > 0$ avoid any potential issue of dividing by zero, but also intuitively makes sense as trying to determine the probability that A occurs given B has occurred, when B could never occur seems illogical at best.

Example:
A die is rolled and the number showing face up is recorded. Given that the number rolled was even, what is the probability that it was a six?

Solution. ✐

4.8 The multiplication rule

Often we want to know $\Pr(A \cap B)$ given that we know $\Pr(A|B)$ and $\Pr(B)$. A simple rearrangement of the conditional probability formula gives us the multiplication rule:

$$\boxed{\Pr(A \cap B) = \Pr(B) \times \Pr(A|B)}$$

Note that it doesn't make any difference if we swap A and B, so the following formula is also true:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A)\,.$$

Examples:

1. Two cards are dealt from a deck of 52 cards. What is the probability that they are both Aces?

   Solution. We know that $\Pr(\text{Ace }) = 4/52 = 1/13$ and

   $$\Pr(\text{2nd Ace } | \text{ given the first Ace }) = 3/51.$$

   So

   $$\begin{aligned}\Pr(\text{ Two Aces }) &= \Pr(\text{Ace })\Pr(\text{2nd Ace } | \text{ given the first Ace }) \\ &= 1/13 \times 3/51 \simeq 0.0045\end{aligned}$$

2. Let K be the event that "Kate arrives on time to her 9am lecture" and M be the event that "the Metro is on time". The probability that Kate arrives on time to her 9am lecture given that the Metro is on time is $\Pr(K|M) = 0.9$. Suppose the probability that the Metro is not on time is $\Pr(M^c) = 0.2$. What is the probability that both Kate *and* the Metro are on time?

   Solution. ✐

4.9 Independent events

Two events A and B are said to be independent if

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)\,.$$

Using the definition of conditional probability we get

$$\Pr(A|B) = \Pr(A)$$

when A and B are independent. You can think of it like this: events are independent when information about one event doesn't change the probability of the other. Example: hair colour and weight are independent (knowing someone's hair colour doesn't affect what they are likely to weigh); hair colour and eye colour are *dependent*.

Example (Roy Meadow and the Sally Clark Case)**:**
A brief outline of the case is as follows.

- In 1999 the solicitor Sally Clark was tried for allegedly murdering her two babies.
- Her elder son Christopher had died at the age of 11 weeks, and her younger son Harry at 8 weeks.
- Prof. Sir Roy Meadow testified that

  $$\Pr(\text{a single cot death}) \simeq \frac{1}{8500}$$

  which he obtained from the observed ratio of live-births to cot deaths in affluent non-smoking families.

- He then went on to say that

$$\Pr(\text{two cot deaths}) \simeq \frac{1}{8500} \times \frac{1}{8500} = \frac{1}{8500^2} = \frac{1}{72,250,000}$$

  or about 1 in 72 million. Based partly on this evidence, the jury returned a 10/2 majority verdict of guilty. There are two problems with this figure.

- Are the events $A =$*first child dies of SIDS* and $B =$*second child dies of SIDS* independent? (SIDS stands for Sudden Infant Death Syndrome.) Genetic or environmental factors might make

$$\Pr(\text{second child dies of SIDS}|\text{first child dies of SIDS})$$

  much higher than 1/8500.

- Aside from its invalidity, figures such as the 1 in 72 million are very easily misinterpreted. Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis-)interpretation is a serious error of logic known as the Prosecutor's Fallacy. The jury needs to weigh up two competing explanations for the babies' deaths: SIDS or murder. Two deaths by SIDS or two murders are each quite unlikely, but one situation has apparently happened in this case. What matters is the relative likelihood of the deaths under each explanation, not just how unlikely they are under one explanation (in this case SIDS, according to the evidence as presented).

Example:
A playing card is drawn from a pack. Let A be the event "an Ace is drawn" and let C be the event "a Club is drawn". Are the events A and C mutually exclusive (i.e. disjoint)? Are they independent?

Solution.

$$\Pr(A \cap C) = \Pr(\text{Ace of clubs}) = \frac{1}{52}$$

but

$$\Pr(A) = \frac{4}{52} \quad \text{and } \Pr(C) = \frac{13}{52}.$$

So $\Pr(A) \times \Pr(C) = 1/52 = \Pr(A \cap C)$. It follows that A and C are **not** mutually exclusive but they are independent.

4.9.1 Mutual independence and pairwise independence

The set of events $A = \{A_1, A_2, \dots, A_n\}$ are *mutually independent events* if *for any subset* $\{A_{i1}, A_{i2}, \dots, A_{ir}\}, r \leq n$ we have

$$\Pr(A_{i1} \cap \cdots \cap A_{ir}) = \Pr(A_{i1}) \times \cdots \times \Pr(A_{ir}) \ .$$

The set of events $A = \{A_1, A_2, \ldots, A_n\}$ are *pairwise independent* if *for any pair* $\{A_{i1}, A_{i2}\}$ we have

$$\Pr(A_{i1} \cap A_{i2}) = \Pr(A_{i1}) \times \Pr(A_{i2}) \ .$$

Note that mutual independence is much stronger that pairwise independence. That is, mutual independence implies pairwise independence but pairwise independence *does not* imply mutual independence.

Example:
Toss two coins and define the events:

- A_1 = head first throw;
- A_2 = head second throw;
- A_3 = head both times or not at all.

Show that these events are pairwise independent but not mutually independent.

Solution.

4.10 The law of total probability

We have seen that for any event B, B and B^c are disjoint, and that

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c) \, .$$

Using the multiplication rule, this gives

$$\boxed{\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)}$$

This is the **Law of Total Probability**.

More generally, a set of events $E_1, E_2, \ldots, E_n$ is said to be an *exhaustive set of events* if $E_1 \cup E_2 \cup \cdots \cup E_n$ is equal to the sample space, i.e. if $\Pr(E_1 \cup E_2 \cup \cdots \cup E_n) = 1$.

A *partition* is a set $E_1, E_2, \ldots, E_n$ of events that are both exhaustive and mutually exclusive.

Suppose that X is some other event and that $\Pr(X|E_j)$ and $\Pr(E_j)$ are known for $j = 1, \ldots, n$. Since $E_1, E_2, \ldots, E_n$ are mutually exclusive $(E_1 \cap X), \ldots, (E_n \cap X)$ are also mutually exclusive.

Since $E_1, E_2, \ldots, E_n$ form an exhaustive set we can therefore write

$$\begin{aligned}\Pr(X) &= \Pr(E_1 \cap X) + \Pr(E_2 \cap X) + \cdots + \Pr(E_n \cap X)\\ &= \Pr(X|E_1)\Pr(E_1) + \Pr(X|E_2)\Pr(E_2) + \cdots + \Pr(X|E_n)\Pr(E_n)\\ &= \sum_{j=1}^{n} \Pr(X|E_j)\Pr(E_j).\end{aligned}$$

Example:

1. A large class consists of 60% female and 40% male students. The females are conscientious and are 90% likely to hand in a homework. Males often think they have better things to do and are 70% likely to hand in the homework. A student is selected at random. What is the probability that they will hand in the homework?

   Solution. ✎

2. 15% of males are left-handed but 12% of the entire population are left-handed. Assuming the population is made up of 50% males and 50% females, what proportion of females are left-handed?

   Solution. ✎

4.11 Tree diagrams

Tree diagrams or probability trees are simple clear ways of presenting probabilistic information. In a tree diagram, experiments are represented by circles (called *nodes*) and the outcomes of the experiments are represented by *branches*. The ends of the branches of the tree are usually known as *terminal nodes*.

The branches are annotated with the probability of the particular outcome. We *multiply* probabilities horizontally along branches ("the multiplication rule") and we *sum* probabilities vertically ("the law of total probability").

Example:
Recall the example of Kate and the Metro. Let K be the event that "Kate arrives on time to her 9am lecture" and M be the event that "the Metro is on time". The probability that Kate arrives on time to her 9am lecture given that the Metro is on time is $\Pr(K|M) = 0.9$. The probability that the Metro is not on time is $\Pr(M^c) = 0.2$. Suppose also that the probability that Kate is not on time given that the Metro is not on time is $\Pr(K^c|M^c) = 0.9$.

Represent this information in a tree diagram and use it to compute the probability that Kate does not arrive on time to her 9am lecture, $\Pr(K^c)$.

Solution.

4.12 Bayes' Theorem

Bayes' Theorem is a way of "reversing the condition" in a conditional probability. It is very widely used in applied statistics, and has given rise to its own branch / philosophy of inference, "Bayesian Inference".

From the definition of conditional probability

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)},$$

substitute in the formula for the multiplication rule on the top line and the law of total probability on the bottom to get:

$$\begin{aligned}\Pr(A|B) &= \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \\ &= \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c)}\end{aligned}$$

This is known as *Bayes' Theorem* after the Reverend Thomas Bayes (1702-1761).

More generally, suppose $E_1, E_2, \ldots, E_n$ is an exhaustive mutually exclusive set (as above). A more general version of Bayes' theorem specifies the probability of each of the possibilities E_k given some other event X:

$$\begin{aligned}\Pr(E_k|X) &= \frac{\Pr(E_k)\Pr(X|E_k)}{\Pr(X)} \\ &= \frac{\Pr(E_k)\Pr(X|E_k)}{\sum_{j=1}^{n}\Pr(E_j)\Pr(X|E_j)},\end{aligned}$$

where E_k is a particular member of $E_1, E_2, \ldots, E_n$.

Examples:

1. A machine is used to manufacture components. When the machine is working properly each component has a probability of 0.01 of turning out to be defective. When the machine has been in use for a certain long period of time there is a probability of 0.1 that a certain part in the machine will have failed. If this happens then the probability of a defective component is increased to 0.5. Find the conditional probability that the machine is working correctly after such a period given that a component produced at that time is examined and found to be defective.

   Solution. Let C be the event "the machine is working correctly" and D be the event "the article is defective". Then

   $$\Pr(C|D) = \frac{\Pr(C)\Pr(D|C)}{\Pr(D)}$$

   and

   $$\Pr(D) = \Pr(C)\Pr(D|C) + \Pr(C^c)\Pr(D|C^c)$$

   so

   $$\Pr(C|D) = \frac{0.9 \times 0.01}{0.9 \times 0.01 + 0.1 \times 0.5} = \frac{0.009}{0.059} = 0.153.$$

Hence also

$$\Pr(C^c|D) = 1 - 0.153 = 0.847.$$

Similarly

$$\Pr(C|D^c) = \frac{0.9 \times 0.99}{0.9 \times 0.99 + 0.1 \times 0.5} = \frac{0.891}{0.941} = 0.947.$$

2. A clinic offers a free test for a rare disease, which affects about 1 in 10,000 people. If you have the disease it has a 98% chance of giving a positive result, and if you don't have the disease, it has only a 1% chance of giving a positive result. You decide to take the test, and find that you test positive. What is the probability that you have the disease?

   Solution.

3. Joe Soap has an examination on Thursday morning. On Wednesday night he is free to choose one (and only one) of the following activities: (a) go to the cinema, (b) go to the pub, (c) stay home and watch TV, (d) stay home and study. The probabilities that he elects these alternatives are 0.14, 0.45, 0.25 and 0.16 respectively. His probabilities of passing the exam given (a), (b), (c) and (d) are 0.4, 0.05, 0.5 and 0.8 respectively. Find

   (i) Pr(Joe passes)

   (ii) Pr(Joe stayed in and watched TV given that he passed)

   Solution. ✐

4.13 Basic combinatorics

Many examples arise where the sample space is finite, with every outcome equally probable. In that case:

$$\Pr(A) = \frac{\text{number of elements in } A}{\text{number of elements in } S}.$$

Computing probabilities boils down to *counting* the number of elements in different sets. In this section of the notes we look at some elementary ideas from *combinatorics* – the mathematics of counting – since we need these ideas to evaluate certain probabilities. Before we start, recall the notation:

$$n! = 1 \times 2 \times \cdots \times n,$$

with $0! = 1$ by definition.

In R, this number is evaluated using the `factorial` function; for example `factorial(3)` returns the value 6.

4.13.1 Sequences

Suppose the set of elements we want to count consists of sequences that are r symbols long and that each position in the sequence can be any one of n symbols. Then the total number of sequences is n^r.

4.13.2 Permutations

Suppose we are choosing a subset of r elements from a bigger set containing n elements. In addition, suppose the r elements are picked out one at a time *without replacement* in such a way that the *order* in which the objects are chosen matters. Such a choice of r objects is called a *permutation*. There are n choices for the first object chosen, then $n-1$ choices for the second given the first, and so on up to $n-r+1$ choices for the r-th object. So

$$\begin{aligned} \begin{array}{c}\text{number of ways of chosing } r \text{ from } n\\ \text{objects without replacement}\end{array} &= n \times (n-1) \times \cdots \times (n-r+1) \\ &= \frac{n!}{(n-r)!}. \end{aligned}$$

This number is denoted ${}^nP_r$ or P_r^n:

$$P_r^n = \frac{n!}{(n-r)!}.$$

4.13.3 Combinations

Now suppose we are picking a subset of r elements from a bigger set of n elements *without replacement* but we want to count the number of ways this can be done where the *order doesn't matter*. We will denote this ${}^nC_r$ or C_r^n, defined as the number of *combinations* of r objects chosen from n. Consider a single one of these combinations. The objects could be ordered in any of $r!$ ways because there are r choices for the object we place first, $r-1$ choices for the object we place second, and so on. This would give us all the permutations corresponding to that particular combination. The same is true of each of the C_r^n combinations and so we have the following formula relating the number of combinations and the number of permutations:

$$P_r^n = C_r^n \times r!$$

Rearranging the above expression, it follows that

$$C_r^n = P_r^n / r! = \frac{n!}{r! \times (n-r)!}.$$

This is also sometimes denoted using the following "brackets" notation:

$$C_r^n = \binom{n}{r}.$$

The number of combinations can also be referred to as the *binomial coefficients*, and are computed in R using the function `choose(n,r)`.

4.13.4 Total number of subsets

Given a set of n elements consider the total number of possible subsets (including the empty set). Each of the n elements can either be included in the subset or not, so there 2^n possible subsets.

Examples:1. There are four nucleotides ("letters") in DNA. How many different DNA sequences of length 10 are there?

2. Suppose that your bank card has just been stolen.
   (a) What is the probability of the thief guessing your 4 digit PIN in one go?
   (b) What is the probability of the thief guessing your 4 digit PIN in one go if he knows that your PIN uses 4 different digits?

3. A coin is tossed 10 times. How many ways can you get 4 heads and 6 tails?

Solution.