1 — A SINGLE RANDOM VARIABLE

Questions involving probability abound in Computer Science:

- What is the probability of the PWF world falling over next week?
- What is the probability of one packet colliding with another in a network?
- What is the probability of an undergraduate not turning up for a lecture?

When addressing such questions there are often added complications: the question may be ill posed or the answer may vary with time. Which undergraduate? What lecture? Is the probability of turning up different on Saturdays?

Let's start with something which appears easy to reason about...

Introduction — Throwing a die

Consider an *experiment* or *trial* which consists of throwing a mathematically ideal die. Such a die is often called a *fair* die or an *unbiased* die. Common sense suggests that:

- The *outcome* of a single throw cannot be predicted.
- The outcome will necessarily be a random integer in the range 1 to 6.
- The six possible outcomes are *equiprobable*, each having a probability of $\frac{1}{6}$.

Without further qualification, serious probabilists would regard this collection of assertions, especially the second, as almost meaningless. Just what is a random integer? Giving proper mathematical rigour to the foundations of probability theory is quite a taxing task.

To illustrate the difficulty, consider probability in a *frequency* sense. Thus a probability of $\frac{1}{6}$ means that, over a long run, one expects to throw a 5 (say) on one-sixth of the occasions that the die is thrown. If the actual proportion of 5s after n throws is $p_5(n)$ it would be nice to say:

$$\lim_{n\to\infty} p_5(n) = \frac{1}{6}$$

Unfortunately this is *utterly bogus mathematics*! This is simply not a proper use of the idea of a limit.

Suppose you frequently play a board game which requires you to 'throw a 6 to start'. Each time you play the game you write down how many throws it took to get your 6. One day you ponder what the long-term average ought to be...

You reason that each of the six possible outcomes should occur with equal frequency in each of the first six throws. You reason further that the required 6 could be in any position from first to sixth in these first six throws and so, on average, you will get the first 6 at throw $3\frac{1}{2}$.

This, too, is nonsense reasoning but you will have to wait a while before the correct answer is derived. Meantime, note a health warning: *relying on intuition can seriously damage your understanding of Probability*!

Relationship to Set Theory — Sample Space — Sample Point

This course will concentrate on solving problems. It is not proposed to venture very far into the foundations of probability theory but some formal discussion is unavoidable.

It *is* mathematically sound to represent the collection of possible outcomes of an experiment as a set. Call this set Ω. In the case of a die:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

The set Ω is referred to as the *sample space* associated with the experiment. Each member of the set is a *sample point.*

Events

There are times when you are concerned not so much with the value which results from a particular throw but with whether this value is in some specific subset of Ω. For example, you might bet on the result being an even number, when the subset of interest is $\{2, 4, 6\}$. Call this subset E:

$$E = \{2, 4, 6\}$$

A set such as E is called an *event.* This term is often used in a slightly unnatural way: for example, you can talk of 'betting on some event'. This clearly doesn't mean that you hope the outcome of a throw is a subset like $\{2, 4, 6\}$ since the outcome can be only a single value. The intended meaning is that you hope the outcome is a *member* of the specified subset.

Two Special Events

Any subset of Ω could be deemed an event. Two special cases are the empty set $\phi = \{\}$ and the full set Ω itself.

The outcome of a throw cannot be a member of the empty set so a bet on ϕ is certain to lose. The outcome of a throw is necessarily a member of Ω so a bet on Ω is certain to win.

Elementary Events

Any event which corresponds to a single sample point is also clearly special. Such an event is called an *elementary event.* For example, the elementary event $\{5\}$ corresponds to the single sample point 5. Note that $\{5\}$ is a *subset* of Ω whereas 5 is a *member* of Ω.

Random Variables — The $\{X{=}r\}$ Notation

A most useful abstraction is the idea of a *random variable.* A naïve view of a random variable is that, rather like a variable in a programming language, it has a *name* and a *value* and whereas the name stays unchanged the value changes.

It is common to use the name X when referring to a single random variable. Suppose X corresponds to the outcome of throwing a die, so the value of X after a particular throw might be 5. It is easy to arrange for a spreadsheet cell to simulate such behaviour. At a given moment the appearance of the cell might be:

$$\begin{array}{c} X \\ \boxed{5} \end{array}$$

It is more usual to write $\{X = 5\}$ and this is an alternative way of representing the elementary event $\{5\}$. In both cases the curly-brackets emphasise that a set is being represented and not an element.

The notation $\{X = 5\}$ can be generalised to $\{X = r\}$. In the context of a die, r would have one of the values 1, 2, 3, 4, 5 or 6. Consider X to be the *name* of the variable and r to be its *value*.

Random Variables — Discrete *versus* Continuous

When X is associated with the outcome of throwing a die, it is an example of a *discrete* random variable, one which can take on only a countable number of different values.

Probability theory extends to *continuous* random variables too. For example, the outcome of throwing a dart at the centre of a dart board can be recorded as the distance of the landing point from the centre. Subject to certain practical constraints, this distance can be any positive real number. Continuous random variables will be left until later.

Combining Events

In discrete examples, any event can be represented as a set and ordinary set notation can be employed. Thus the combined event $\{X = 2\} \cup \{X = 4\} \cup \{X = 6\}$ is equivalent to the event $\{2, 4, 6\}$ and if $E = \{2, 4, 6\}$ then the complementary event $\bar{E} = \{1, 3, 5\}$.

Events are said to be *exclusive* if the associated sets are disjoint. Thus $\{1, 4\}$ and $\{2, 6\}$ are exclusive events.

Events are said to be *exhaustive* if the union of the associated sets incorporates every possible sample point. Thus $\{1, 2, 3, 4\}$ and $\{3, 4, 5, 6\}$ are exhaustive.

Note that the two events E and $\bar{E}$ are exclusive and exhaustive as are Ω and ϕ.

The Capital-P Notation

As shorthand for writing 'the probability of...' it is customary to write capital-P followed by a pair of round brackets. The brackets enclose the event whose probability is of interest. Capital-P is best introduced by means of some examples:

$\mathrm{P}(\{2, 4, 6\})$	means	The probability of the event $\{2, 4, 6\}$.
$\mathrm{P}(E)$	means	The probability of the event E.
$\mathrm{P}(\{5\})$	means	The probability of the elementary event $\{5\}$.
$\mathrm{P}(\{X = r\})$	means	The probability of the random variable X being r.

As clarified earlier, 'the probability of the event E' really means 'the probability that the outcome is a member of E'.

If the probability is known, P(*some event*) can appear on the left-hand side of an equation:

$\mathrm{P}(\{2, 4, 6\}) = \frac{1}{2}$	means	The probability of the event $\{2, 4, 6\}$ is $\frac{1}{2}$.
$\mathrm{P}(E) = \frac{1}{2}$	means	The probability of the event E is $\frac{1}{2}$.
$\mathrm{P}(\{5\}) = \frac{1}{6}$	means	The probability of the elementary event $\{5\}$ is $\frac{1}{6}$.
$\mathrm{P}(\{X = 5\}) = \frac{1}{6}$	means	The probability of the random variable X being 5 is $\frac{1}{6}$.

Common sense suggests that $\mathrm{P}(\{2, 4, 6\}) = \frac{1}{2}$ and this is shown formally on page 1.5.

Axioms of Probability

Taking Ω as sample space and A as an event (a subset of Ω), three assertions can be taken as axioms:

I $\mathrm{P}(A) \geqslant 0$. The probability of any event is positive.

II $\mathrm{P}(\Omega) = 1$. The probability of certainty is 1.

III If $A_1, A_2, A_3 \ldots$ are exclusive (disjoint) events (such that $A_i \cap A_j = \phi$ whenever $i \neq j$) then $\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup \ldots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) \ldots$

Axiom III is known as the Addition Rule of Probability.

The Empty Set Theorem — P(ϕ)=0

The result $\mathrm{P}(\phi) = 0$, although self-evident, (the probability of an impossible event is 0) is nevertheless a theorem whose proof follows from the axioms...

Given that $\Omega = \Omega \cup \phi$, it follows that:

$$\mathrm{P}(\Omega) = \mathrm{P}(\Omega \cup \phi)$$

From axiom II, the left-hand side $\mathrm{P}(\Omega) = 1$ and, from axiom III, the right-hand side $\mathrm{P}(\Omega \cup \phi) = \mathrm{P}(\Omega) + \mathrm{P}(\phi)$ since Ω and ϕ are exclusive. Accordingly, $\mathrm{P}(\Omega) + \mathrm{P}(\phi) = 1$ which, given that P(Ω)=1, immediately ensures that $\mathrm{P}(\phi) = 0$.

The Summation of Elementary Events Theorem

If event A is the set of sample points $\{s_1, s_2, \ldots, s_n\}$ then:

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(\{s_i\})$$

In words: the probability of event A is the sum of the probabilities of the elementary events whose union is A.

Any non-empty event is the union of certain elementary events and elementary events are necessarily exclusive. In particular $A = \{s_1\} \cup \{s_2\} \cup \ldots \cup \{s_n\}$ so the theorem is just a special case of axiom III.

The Complementary Event Theorem

Let event $\bar{A}$ be the complement of event A (strictly the complement with respect to Ω). The complementary event theorem states that:

$$\mathrm{P}(A) = 1 - \mathrm{P}(\bar{A})$$

Given that $A \cup \bar{A} = \Omega$ it follows that $\mathrm{P}(A \cup \bar{A}) = \mathrm{P}(\Omega) = 1$.

Now A and $\bar{A}$ are exclusive, so $\mathrm{P}(A \cup \bar{A}) = \mathrm{P}(A) + \mathrm{P}(\bar{A})$ by axiom III.

Accordingly, $\mathrm{P}(A) + \mathrm{P}(\bar{A}) = 1$. This proves the theorem.

Observations on the Axioms and Theorems

Being axioms, the assertions are not subject to proof but they are clearly reasonable. Taking probability in the frequency sense, it is clear that the proportion of fives in a long run of throws of a die cannot be negative. It is equally clear that throwing *some* value in the range one to six is certain.

It seems simply common sense that the probability of an impossible event is 0 but taking the probability of certainty to be 1 is merely convenient. It would be possible to take it as axiomatic that the probability of certainty were 100 but this would lead to tiresome factors of 100 appearing in calculations.

Event $E = \{2, 4, 6\}$ is the union of the three elementary events $\{2\}$, $\{4\}$ and $\{6\}$. The probability of each of these elementary events is $\frac{1}{6}$ so the sum of the probabilities is $\frac{1}{2}$.

Thus the demonstration that $\mathrm{P}(E) = \frac{1}{2}$ stems directly from the summation of elementary events theorem.

Further use of Sets — Event Space

When the sample space is as simple as $\Omega = \{1, 2, 3, 4, 5, 6\}$, the power set (the set of all possible subsets) of Ω contains every possible event. The power set is an example of an *event space* associated with a sample space. (For a fuller understanding of event space see exercise 9.)

Consider the two particular events from the event space:

$$E = \{2, 4, 6\} \qquad \text{and} \qquad S = \{1, 4\}$$

With reference to throwing a die, event E is the outcome of throwing an even number and event S is the outcome of throwing a perfect square.

With the exception of the event ϕ, any event is the union of one or more elementary events. In the case of a fair die, the six elementary events are equiprobable (the probability is $\frac{1}{6}$ in each case).

Noting these points and the summation of elementary events theorem:

$$\mathrm{P}(E) = \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{3}{6}$$

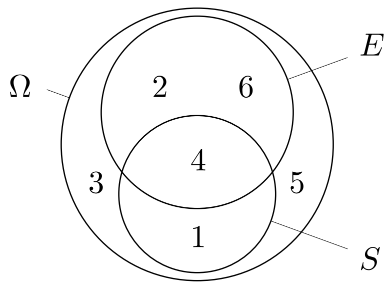
$$\mathrm{P}(S) = \mathrm{P}(\{1\}) + \mathrm{P}(\{4\}) = \frac{2}{6}$$

$$\mathrm{P}(E \cap S) = \mathrm{P}(\{4\}) = \frac{1}{6}$$

$$\mathrm{P}(E \cup S) = \mathrm{P}(\{1, 2, 4, 6\}) = \mathrm{P}(\{1\}) + \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{4}{6}$$

A Venn Diagram Illustration

A Venn diagram neatly illustrates the foregoing results:



Clearly $E \cap S$ is the elementary event $\{4\}$ and the probability of this event is $\frac{1}{6}$.

By contrast $E \cup S$ is the union of four elementary events so its probability is $\frac{4}{6}$.

The number of members in a set A is the *cardinality* of the set and is written as $|A|$. With a die the elementary events are equiprobable, so it is easy to see that the probability of any event $A \subseteq \Omega$ is:

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}$$

With a die $|\Omega| = 6$. In the examples above, $|E| = 3$, $|S| = 2$, $|E \cap S| = 1$ and $|E \cup S| = 4$ so the probabilities of these events are $\frac{3}{6}$, $\frac{2}{6}$, $\frac{1}{6}$ and $\frac{4}{6}$ respectively.

Probability as Area

In simple cases, the areas of the different regions in a Venn diagram representation can be made to reflect probabilities. In the figure, the circle representing event Ω should be taken to have unit area to reflect that $\mathrm{P}(\Omega) = 1$ (certainty) by axiom II. The area of the circle representing event E is constrained to be one half to reflect that $\mathrm{P}(E) = \frac{1}{2}$ and the area of the circle representing event S is constrained to be one third because $\mathrm{P}(S) = \frac{1}{3}$.

In principle, there could be further subdivisions which gave each elementary event an area of one sixth.

What if the Elementary Events are not Equiprobable?

The axioms of probability apply generally and there is no requirement for the elementary events to be equiprobable. With a weighted die, for example, there may be a different probability of throwing each of the six values. The probability of an event such as $\{2, 4, 6\}$ would still be the sum of the probabilities of the elementary events but this would not necessarily be $\frac{1}{2}$.

Clearly additional information is required for a full specification of the properties of a given die. A simple approach is to present the probability of each elementary event and, from these values, the probabilities of any event in the power set of events can be computed.

Probability Space

From any (straightforward) sample space, the associated power set gives an event space. When this information is augmented by the probabilities of each individual event the whole is referred to as the associated *probability space*.

Clearly there is redundant information in a fully-specified probability space since specifying the probabilities of the elementary events alone provides sufficient information to calculate the probabilities of all the other possible events.

Sometimes just the probabilities of certain non-elementary events may be provided and, in appropriate circumstances, it is possible to derive either the probabilities of the associated elementary events or the probabilities of other non-elementary events. (See exercise 6.)

Representing the Probabilities of Elementary Events

Even when the elementary events are not equiprobable, the areas of the different regions in a Venn diagram representation can still be made to correspond to probabilities in simple cases.

In more ambitious cases there are alternative ways of specifying the separate probabilities of each elementary event...

For each $r \in \Omega$ the value of $\mathrm{P}(\{X = r\})$ must be specified. One way to achieve this is to draw up a table. Consider a rather carefully constructed die which is biased so that the probability of throwing r is proportional to r. An appropriate table of probabilities of the elementary events is:

$r \longrightarrow$

X	1	2	3	4	5	6
$\mathrm{P}(\{X = r\})$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

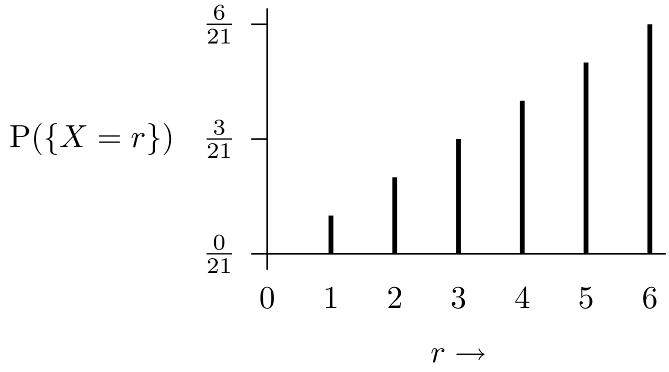
It is well worth developing an instinctive reaction to meeting such a table: *check that the sum of the probabilities of the elementary events is unity.* This always has to be the case since $\mathrm{P}(\Omega) = 1$.

$\mathrm{P}(\{X = r\})$ is a function of r and can be described in ordinary mathematical notation:

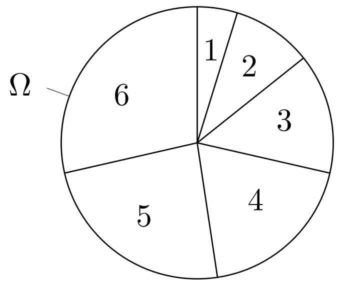
$$\mathrm{P}(\{X = r\}) = \begin{cases} \dfrac{r}{21}, & \text{if } r \in \mathbb{N} \wedge 1 \leqslant r \leqslant 6 \\ 0, & \text{otherwise} \end{cases}$$

Such a function is sometimes known as a *density function* but this term is more commonly used for continuous random variables.

The function can be represented graphically:



A pie chart is another useful way of representing the different probabilities:



The pie as a whole has unit area and the area of each slice is equal to the probability of the elementary event which labels the slice.

The pie chart also suggests a way of constructing a practical device which will come up with the required probabilities. The chart could be regarded rather like the card on a magnetic compass but it should be fitted with a needle which isn't magnetised. An experiment consists of spinning the needle and allowing it to come to rest. The label on the slice at which the head end of the needle then points is assigned to the random variable X.

It would be very hard to construct a die which is biased in the corresponding way!

The Summation of Elementary Events Theorem revisited

Suppose that, using the biased die, the events E and S are again:

$$E = \{2, 4, 6\} \qquad \text{and} \qquad S = \{1, 4\}$$

The examples given on page 1.5 can be reworked using the new probabilities for the elementary events:

$$\mathrm{P}(E) = \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{12}{21}$$

$$\mathrm{P}(S) = \mathrm{P}(\{1\}) + \mathrm{P}(\{4\}) = \frac{5}{21}$$

$$\mathrm{P}(E \cap S) = \mathrm{P}(\{4\}) = \frac{4}{21}$$

$$\mathrm{P}(E \cup S) = \mathrm{P}(\{1,2,4,6\}) = \mathrm{P}(\{1\}) + \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{13}{21}$$

Given the particular biased die and the specified events E and S, the four results just obtained happen to be related as follows:

$$\frac{13}{21} = \frac{12}{21} + \frac{5}{21} - \frac{4}{21}$$

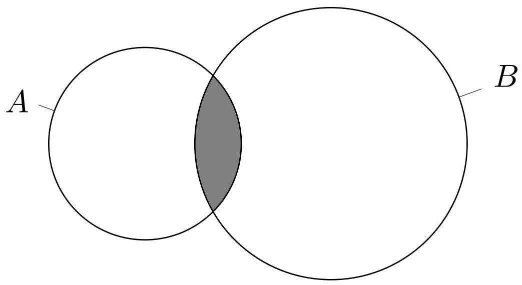
Accordingly, in these very particular circumstances:

$$\mathrm{P}(E \cup S) = \mathrm{P}(E) + \mathrm{P}(S) - \mathrm{P}(E \cap S)$$

This can be shown to be true generally...

The Inclusion-Exclusion Theorem

Suppose A and B are two arbitrary events in some sample space. Consider the following Venn diagram in which area corresponds to probability:



The area of $A \cup B$ includes the area of A and the area of B but is not the sum of the areas of A and B since this includes the shaded region representing $A \cap B$ twice. In consequence the sum of the areas of A and B is greater than the area of $A \cup B$ by an amount equal to $A \cap B$. Translating areas into probabilities gives:

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

The sum of the terms $\mathrm{P}(A)$ and $\mathrm{P}(B)$ on the right-hand side *includes* the shaded region twice and the third term *excludes* the shaded region once. The relationship is, accordingly, known as the inclusion-exclusion theorem and applies generally. See exercises 10 and 11.

If A and B are exclusive events $A \cap B = \phi$ and the theorem conforms with axiom III...

$$\text{If } A \cap B = \phi \qquad \text{then} \qquad \mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$$

Conditional Probability

The notation of which $\mathrm{P}(\{X = 5\})$ is an example readily extends to forms such as:

$$\mathrm{P}(\{X \neq 4\}) \qquad \text{or} \qquad \mathrm{P}(\{X > 4\})$$

From the complementary event theorem it quickly follows that:

$$\mathrm{P}(\{X \neq 4\}) = 1 - \mathrm{P}(\{X = 4\})$$

The second example, $\mathrm{P}(\{X > 4\})$, is known as a *tail probability.* In this case the *upper* tail-end values 5 and 6 are referred to and assuming the discussion is still about a die:

$$\mathrm{P}(\{X > 4\}) = \mathrm{P}(\{5, 6\})$$

A more important notion of *conditional probability* occurs when asking a question of the form 'What is the probability that a perfect square has been thrown *given that* an even number has been thrown?'

The notation $\big|$ is used to mean 'given' and the question just posed is more formally written:

$$\mathrm{P}(\{1, 4\} \,\big|\, \{2, 4, 6\})$$

In general: $\mathrm{P}(B \,|\, A)$ means The probability of event B given event A.

By consideration of probability as area, it is easy to show that:

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

Refer to the Venn diagram on the previous page. Given event A, any sample point has to be in the circle representing A. For event B to occur, *given* event A, the sample point has to be in the shaded region.

The probability on the left-hand side, $\mathrm{P}(B \,|\, A)$, is therefore the ratio of the area of the shaded region to the area of the circle representing A. This is the right-hand side.

Mapping

The outcome of a throw of a die is necessarily a number. This makes it possible to make informal assertions such as 'on average the outcome is $3\frac{1}{2}$'.

Very often an outcome is not numerical, for example, 'the horse that wins the 3 o'clock at Newmarket'. The $\mathrm{P}(\{X = r\})$ notation extends readily to non-numerical values for r. Consider a particular case...

Suppose a bag contains six balls, one red, two white and three blue and, on picking out a ball at random, each of the six balls is equally likely to be selected. One can readily imagine betting on the colour of the ball that is picked out.

To analyse this example, the sample space $\Omega = \{\text{red}, \text{white}, \text{blue}\}$ and for each $r \in \Omega$ (that is for each of the three elementary events) the value of $\mathrm{P}(X = r)$ can be tabulated:

$$
\begin{array}{r|c|c|c|}
 & & r \longrightarrow & \\
X & \text{red} & \text{white} & \text{blue} \\
\cline{2-4}
\mathrm{P}(\{X = r\}) & \frac{1}{6} & \frac{2}{6} & \frac{3}{6} \\
\cline{2-4}
\end{array}
$$

Clearly the probabilities can equally be represented by a graph or by a pie chart. Questions like what is the probability of picking out a ball that is either white or blue can now readily be answered.

With non-numerical values it is normally meaningless to ask about the average outcome. This is not something to worry about immediately but many techniques in probability theory assume that any outcome *is* a number. In particular, many formulae in probability theory involve summing over r and, with discrete random variables, it is usually assumed that each $r \in \mathbb{N}$.

There is a simple way to accommodate the requirement for r to be a number. This is to map the possible outcomes onto the set of numbers $\{0, 1, 2, \ldots\}$. The mapping is often arbitrary and in the present case assigning red=0, white=1 and blue=2 is as good as any other mapping. The table would then become:

$$
\begin{array}{r|c|c|c|}
 & & r \longrightarrow & \\
X & 0 & 1 & 2 \\
\cline{2-4}
\mathrm{P}(\{X = r\}) & \frac{1}{6} & \frac{2}{6} & \frac{3}{6} \\
\cline{2-4}
\end{array}
$$

Note that r ranges from 0 upwards (and not 1 upwards). Using 0 as the starting value of r is assumed in many probability formulae so it is unfortunate that the values indicated on a conventional die run from 1 to 6 and not 0 to 5.

It would be altogether eccentric to insist that the faces of any die used in this course carried the values 0 to 5 and an alternative ploy will sometimes be used. It will be assumed that the experiment of throwing a die can yield *seven* different elementary events. These are the members of the set $\{0, 1, 2, 3, 4, 5, 6\}$ with the added proviso that $\mathrm{P}(\{X = 0\}) = 0$.

Glossary

A number of technical terms have been introduced. Those which relate to probability (as distinct from set theory) are, in order of first appearance:

experiment	sample space	exclusive
trial	sample point	exhaustive
fair	event	event space
unbiased	elementary event	probability space
outcome	random variable	density function
equiprobable	discrete	tail probability
frequency	continuous	conditional probability

Check that you understand each of them.

Exercises — I

Whenever possible (and appropriate) *work in fractions* and not in decimals. Show the probability of the event $\{2, 3, 5, 6\}$ when a fair die is thrown as $\frac{2}{3}$ and *not* as 0.6667 or some such.

1. An insurance company is interested in the age distribution of couples. Let A be the event 'husband is older than 40', B be the event 'husband is older than wife' and C be the event 'wife is older than 40'. Describe the following events in plain English:

   (a) $A \cap B \cap C$ $\quad$ (b) $A - A \cap B$ $\quad$ (c) $A \cap \bar{B} \cap C$

2. Given sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$, write down:

   (a) Two events which are exclusive but not exhaustive.

   (b) Two events which are exhaustive but not exclusive.

   (c) Two events which are neither exclusive nor exhaustive.

3. Consider a random variable X whose value represents the outcome of throwing the biased die described on pages 1.7 and 1.8. Determine the following probabilities:

   (a) $\mathrm{P}(\{X = 1\} \cup \{X = 2\})$.

   (b) $\mathrm{P}(\{X < 3\})$.

   (c) $\mathrm{P}(\{X \leqslant 2\})$.

4. Among the digits 1, 2, 3, 4, 5 first one is chosen and then a second selection is made among the four remaining digits. List the 20 possible elementary events. Assume the 20 are equiprobable then find the probability of the following events:

   (a) An odd digit is selected first time.

   (b) An odd digit is selected second time.

   (c) An odd digit is selected both times.

5. The four digits 1, 2, 3, 4 are arranged in some random order. List the 24 possible elementary events. Assume the 24 are equiprobable. Let A_i be the event 'digit i appears in its natural place' (note $i \in \{1, 2, 3, 4\}$). Verify the formula:

$$\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2)$$

6. [From Grimmett and Welsh] Let A, B and C be three events such that:

$$\mathrm{P}(A) = \tfrac{5}{10} \qquad \mathrm{P}(B) = \tfrac{7}{10} \qquad \mathrm{P}(C) = \tfrac{6}{10}$$

$$\mathrm{P}(A \cap B) = \tfrac{3}{10} \qquad \mathrm{P}(B \cap C) = \tfrac{4}{10} \qquad \mathrm{P}(C \cap A) = \tfrac{2}{10}$$

$$\mathrm{P}(A \cap B \cap C) = \tfrac{1}{10}$$

   By drawing a Venn diagram or otherwise, find the probability that exactly two of the events A, B and C occur.

7. A card is selected at random from a normal deck of 52 playing cards. The elementary events can be represented by the set $\{1\spadesuit, \ldots, \text{K}\clubsuit\}$ and these events may be mapped onto the set $\{1, \ldots, 52\}$. Assume that the 52 elementary events are equiprobable.

   Let A be the elementary event $4\diamondsuit$, B be the event 'any ace' and C be the event 'any diamond'. Determine the following probabilities:

   $\mathrm{P}(A) \quad \mathrm{P}(\bar{A}) \quad \mathrm{P}(B) \quad \mathrm{P}(C) \quad \mathrm{P}(B \cup C) \quad \mathrm{P}(B \cap C) \quad \mathrm{P}(A \,|\, B) \quad \mathrm{P}(B \,|\, C)$

8. You overhear a couple talking about a visit to an exclusive girls-only school where they have recently been entertained as prospective parents. You gather that they have two children (and no more are expected). What is the probability that both children are girls?

   You may assume:

   (a) Boys and girls are born equiprobably.

   (b) The sexes of the different children of the same parents are independent.

   The first assumption is only slightly erroneous (in Britain the ratio, at birth, of boys to girls is about 515:485). The second assumption is the subject of debate but it is very widely accepted.

9. Grimmett and Welsh define an event space associated with a sample space Ω as any set of subsets of Ω that satisfies certain constraints. For $\mathcal{E}$ to be an event space associated with sample space Ω, the constraints are:

   (i) $\mathcal{E}$ must be non-empty.

   (ii) If $A \in \mathcal{E}$ then $\bar{A} \in \mathcal{E}$.

   (iii) If $A_1, A_2, A_3 \ldots \in \mathcal{E}$ then $A_1 \cup A_2 \cup A_3 \cup \ldots \in \mathcal{E}$.

   Taking $\Omega = \{1,2,3,4,5,6\}$, consider the smallest associated event space $\mathcal{E}$ that includes the event $\{1\}$. Clearly, $\{2,3,4,5,6\} \in \mathcal{E}$ by (ii). Further, by (iii), $\{1,2,3,4,5,6\} = \Omega \in \mathcal{E}$. By ($ii$) again, $\phi \in \mathcal{E}$, so $\mathcal{E} = \{\phi, \{1\}, \{2,3,4,5,6\}, \Omega\}$.

   By a similar argument, the smallest event space $\mathcal{E}$ that includes the arbitrary event A is $\mathcal{E} = \{\phi, A, \bar{A}, \Omega\}$. One notes that ϕ and Ω are necessarily members of any event space associated with sample space Ω.

   An event space $\mathcal{E}$ is said to be 'closed under the operations of taking complements and countable unions' and as a special case the power set of any sample space is an event space.

   Given sample space $\Omega = \{1,2,3,4,5,6\}$, determine the smallest possible event space $\mathcal{E}$ in each of the following cases:

   (a) The events $\{1\}$ and $\{2\}$ are members of $\mathcal{E}$.

   (b) The events $\{1,2\}$ and $\{3,4\}$ are members of $\mathcal{E}$.

   (c) The events $\{1,2\}$ and $\{1,3\}$ are members of $\mathcal{E}$.

   (d) The events A and B are members of $\mathcal{E}$.

10. If A_1 and A_2 are two arbitrary events in some sample space, the inclusion-exclusion theorem, discussed on page 1.9, can be expressed as:

$$\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2)$$

By using a three-set Venn diagram or otherwise, show that for three arbitrary events A_1, A_2 and A_3 in some sample space the inclusion-exclusion theorem can be expressed as:

$$\begin{aligned}\mathrm{P}(A_1 \cup A_2 \cup A_3) = {} & \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}(A_3) \\ & - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \mathrm{P}(A_2 \cap A_3) \\ & + \mathrm{P}(A_1 \cap A_2 \cap A_3)\end{aligned}$$

11. By considering the effect of introducing a fourth event A_4, state the inclusion-exclusion theorem as it applies to four arbitrary events A_1, A_2, A_3 and A_4 in the same sample space:

$$\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4) = \quad ?$$