1

Events and their probabilities

Summary. Any experiment involving randomness can be modelled as a probability space. Such a space comprises a set Ω of possible outcomes of the experiment, a set $\mathcal{F}$ of events, and a probability measure $\mathbb{P}$. The definition and basic properties of a probability space are explored, and the concepts of conditional probability and independence are introduced. Many examples involving modelling and calculation are included.

1.1 Introduction

Much of our life is based on the belief that the future is largely unpredictable. For example, games of chance such as dice or roulette would have few adherents if their outcomes were known in advance. We express this belief in chance behaviour by the use of words such as 'random' or 'probability', and we seek, by way of gaming and other experience, to assign quantitative as well as qualitative meanings to such usages. Our main acquaintance with statements about probability relies on a wealth of concepts, some more reasonable than others. A mathematical theory of probability will incorporate those concepts of chance which are expressed and implicit in common rational understanding. Such a theory will formalize these concepts as a collection of axioms, which should lead directly to conclusions in agreement with practical experimentation. This chapter contains the essential ingredients of this construction.

1.2 Events as sets

Many everyday statements take the form 'the chance (or probability) of A is p', where A is some event (such as 'the sun shining tomorrow', 'Cambridge winning the Boat Race', ...) and p is a number or adjective describing quantity (such as 'one-eighth', 'low', ...). The occurrence or non-occurrence of A depends upon the chain of circumstances involved. This chain is called an *experiment* or *trial*; the result of an experiment is called its *outcome*. In general, we cannot predict with certainty the outcome of an experiment in advance of its completion; we can only list the collection of possible outcomes.

(1) Definition. The set of all possible outcomes of an experiment is called the **sample space** and is denoted by Ω.

(2) Example. A coin is tossed. There are two possible outcomes, heads (denoted by H) and tails (denoted by T), so that $\Omega = \{\mathrm{H}, \mathrm{T}\}$. We may be interested in the possible occurrences of the following events:

(a) the outcome is a head;

(b) the outcome is either a head or a tail;

(c) the outcome is both a head and a tail (this seems very unlikely to occur);

(d) the outcome is not a head. ●

(3) Example. A die is thrown once. There are six possible outcomes depending on which of the numbers 1, 2, 3, 4, 5, or 6 is uppermost. Thus $\Omega = \{1, 2, 3, 4, 5, 6\}$. We may be interested in the following events:

(a) the outcome is the number 1;

(b) the outcome is an even number;

(c) the outcome is even but does not exceed 3;

(d) the outcome is not even. ●

We see immediately that each of the events of these examples can be specified as a subset A of the appropriate sample space Ω. In the first example they can be rewritten as

(a) $A = \{\mathrm{H}\}$, (b) $A = \{\mathrm{H}\} \cup \{\mathrm{T}\}$,

(c) $A = \{\mathrm{H}\} \cap \{\mathrm{T}\}$, (d) $A = \{\mathrm{H}\}^{\mathrm{c}}$,

whilst those of the second example become

(a) $A = \{1\}$, (b) $A = \{2, 4, 6\}$,

(c) $A = \{2, 4, 6\} \cap \{1, 2, 3\}$, (d) $A = \{2, 4, 6\}^{\mathrm{c}}$.

The *complement* of a subset A of Ω is denoted here and subsequently by A^{c}; from now on, subsets of Ω containing a single member, such as $\{\mathrm{H}\}$, will usually be written without the containing braces.

Henceforth we think of *events* as subsets of the sample space Ω. Whenever A and B are events in which we are interested, then we can reasonably concern ourselves also with the events $A \cup B$, $A \cap B$, and A^{c}, representing 'A or B', 'A and B', and 'not A' respectively. Events A and B are called *disjoint* if their intersection is the empty set $\varnothing$; $\varnothing$ is called the *impossible event*. The set Ω is called the *certain event*, since some member of Ω will certainly occur.

Thus events are subsets of Ω, but need all the subsets of Ω be events? The answer is *no*, but some of the reasons for this are too difficult to be discussed here. It suffices for us to think of the collection of events as a subcollection $\mathcal{F}$ of the set of all subsets of Ω. This subcollection should have certain properties in accordance with the earlier discussion:

(a) if $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$;

(b) if $A \in \mathcal{F}$ then $A^{\mathrm{c}} \in \mathcal{F}$;

(c) the empty set $\varnothing$ belongs to $\mathcal{F}$.

Any collection $\mathcal{F}$ of subsets of Ω which satisfies these three conditions is called a *field*. It follows from the properties of a field $\mathcal{F}$ that

$$\text{if} \quad A_1, A_2, \ldots, A_n \in \mathcal{F} \quad \text{then} \quad \bigcup_{i=1}^{n} A_i \in \mathcal{F};$$

Typical notation	Set jargon	Probability jargon
Ω	Collection of objects	Sample space
ω	Member of Ω	Elementary event, outcome
A	Subset of Ω	Event that some outcome in A occurs
A^{c}	Complement of A	Event that no outcome in A occurs
$A \cap B$	Intersection	Both A and B
$A \cup B$	Union	Either A or B or both
$A \setminus B$	Difference	A, but not B
$A \triangle B$	Symmetric difference	Either A or B, but not both
$A \subseteq B$	Inclusion	If A, then B
$\varnothing$	Empty set	Impossible event
Ω	Whole space	Certain event

Table 1.1. The jargon of set theory and probability theory.

that is to say, $\mathcal{F}$ is closed under finite unions and hence under finite intersections also (see Problem (1.8.3)). This is fine when Ω is a finite set, but we require slightly more to deal with the common situation when Ω is infinite, as the following example indicates.

(4) Example. A coin is tossed repeatedly until the first head turns up; we are concerned with the number of tosses before this happens. The set of all possible outcomes is the set $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$, where ω_i denotes the outcome when the first $i - 1$ tosses are tails and the ith toss is a head. We may seek to assign a probability to the event A, that the first head occurs after an even number of tosses, that is, $A = \{\omega_2, \omega_4, \omega_6, \dots\}$. This is an infinite countable union of members of Ω and we require that such a set belong to $\mathcal{F}$ in order that we can discuss its probability. ●

Thus we also require that the collection of events be closed under the operation of taking countable unions. Any collection of subsets of Ω with these properties is called a *σ-field*.

(5) Definition. A collection $\mathcal{F}$ of subsets of Ω is called a **σ-field** if it satisfies the following conditions:

(a) $\varnothing \in \mathcal{F}$;

(b) if $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$;

(c) if $A \in \mathcal{F}$ then $A^{c} \in \mathcal{F}$.

It follows from Problem (1.8.3) that σ-fields are closed under the operation of taking countable intersections. Here are some examples of σ-fields.

(6) Example. The smallest σ-field associated with Ω is the collection $\mathcal{F} = \{\varnothing, \Omega\}$. ●

(7) Example. If A is any subset of Ω then $\mathcal{F} = \{\varnothing, A, A^{c}, \Omega\}$ is a σ-field. ●

(8) Example. The *power set* of Ω, which is written $\{0, 1\}^{\Omega}$ and contains all subsets of Ω, is obviously a σ-field. For reasons beyond the scope of this book, when Ω is infinite, its power set is too large a collection for probabilities to be assigned reasonably to all its members. ●

To recapitulate, with any experiment we may associate a pair $(\Omega, \mathcal{F})$, where Ω is the set of all possible outcomes or *elementary events* and $\mathcal{F}$ is a σ-field of subsets of Ω which contains all the events in whose occurrences we may be interested; henceforth, to call a set A an *event* is equivalent to asserting that A belongs to the σ-field in question. We usually translate statements about combinations of events into set-theoretic jargon; for example, the event that both A and B occur is written as $A \cap B$. Table 1.1 is a translation chart.

Exercises for Section 1.2

1. Let $\{A_i : i \in I\}$ be a collection of sets. Prove 'De Morgan's Laws'†:

$$\left(\bigcup_i A_i\right)^{c} = \bigcap_i A_i^{c}, \qquad \left(\bigcap_i A_i\right)^{c} = \bigcup_i A_i^{c}.$$

2. Let A and B belong to some σ-field $\mathcal{F}$. Show that $\mathcal{F}$ contains the sets $A \cap B$, $A \setminus B$, and $A \triangle B$.

3. A conventional knock-out tournament (such as that at Wimbledon) begins with 2^n competitors and has n rounds. There are no play-offs for the positions $2, 3, \ldots, 2^n - 1$, and the initial table of draws is specified. Give a concise description of the sample space of all possible outcomes.

4. Let $\mathcal{F}$ be a σ-field of subsets of Ω and suppose that $B \in \mathcal{F}$. Show that $\mathcal{G} = \{A \cap B : A \in \mathcal{F}\}$ is a σ-field of subsets of B.

5. Which of the following are identically true? For those that are not, say when they are true.

(a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;

(b) $A \cap (B \cap C) = (A \cap B) \cap C$;

(c) $(A \cup B) \cap C = A \cup (B \cap C)$;

(d) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

1.3 Probability

We wish to be able to discuss the likelihoods of the occurrences of events. Suppose that we repeat an experiment a large number N of times, keeping the initial conditions as equal as possible, and suppose that A is some event which may or may not occur on each repetition. Our experience of most scientific experimentation is that the proportion of times that A occurs settles down to some value as N becomes larger and larger; that is to say, writing $N(A)$ for the number of occurrences of A in the N trials, the ratio $N(A)/N$ appears to converge to a constant limit as N increases. We can think of the ultimate value of this ratio as being the probability $\mathbb{P}(A)$ that A occurs on any particular trial‡; it may happen that the empirical ratio does not behave in a coherent manner and our intuition fails us at this level, but we shall not discuss this here. In practice, N may be taken to be large but finite, and the ratio $N(A)/N$ may be taken as an approximation to $\mathbb{P}(A)$. Clearly, the ratio is a number between zero and one; if $A = \varnothing$ then $N(\varnothing) = 0$ and the ratio is 0, whilst if $A = \Omega$ then $N(\Omega) = N$ and the

†Augustus De Morgan is well known for having given the first clear statement of the principle of mathematical induction. He applauded probability theory with the words: "The tendency of our study is to substitute the satisfaction of mental exercise for the pernicious enjoyment of an immoral stimulus".

‡This superficial discussion of probabilities is inadequate in many ways; questioning readers may care to discuss the philosophical and empirical aspects of the subject amongst themselves (see Appendix III).

ratio is 1. Furthermore, suppose that A and B are two disjoint events, each of which may or may not occur at each trial. Then

$$N(A \cup B) = N(A) + N(B)$$

and so the ratio $N(A \cup B)/N$ is the sum of the two ratios $N(A)/N$ and $N(B)/N$. We now think of these ratios as representing the probabilities of the appropriate events. The above relations become

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \qquad \mathbb{P}(\varnothing) = 0, \qquad \mathbb{P}(\Omega) = 1.$$

This discussion suggests that the probability function $\mathbb{P}$ should be *finitely additive*, which is to say that

$$\text{if } A_1, A_2, \dots, A_n \text{ are disjoint events, then } \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i);$$

a glance at Example (1.2.4) suggests the more extensive property that $\mathbb{P}$ be *countably additive*, in that the corresponding property should hold for countable collections $A_1, A_2, \dots$ of disjoint events.

These relations are sufficient to specify the desirable properties of a probability function $\mathbb{P}$ applied to the set of events. Any such assignment of likelihoods to the members of $\mathcal{F}$ is called a *probability measure*. Some individuals refer informally to $\mathbb{P}$ as a 'probability distribution', especially when the sample space is finite or countably infinite; this practice is best avoided since the term 'probability distribution' is reserved for another purpose to be encountered in Chapter 2.

(1) Definition. A **probability measure** $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying

(a) $\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\Omega) = 1$;

(b) if $A_1, A_2, \dots$ is a collection of disjoint members of $\mathcal{F}$, in that $A_i \cap A_j = \varnothing$ for all pairs i, j satisfying $i \neq j$, then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$, comprising a set Ω, a σ-field $\mathcal{F}$ of subsets of Ω, and a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, is called a **probability space**.

A probability measure is a special example of what is called a *measure* on the pair $(\Omega, \mathcal{F})$. A measure is a function $\mu : \mathcal{F} \to [0, \infty)$ satisfying $\mu(\varnothing) = 0$ together with (b) above. A measure μ is a probability measure if $\mu(\Omega) = 1$.

We can associate a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with any experiment, and all questions associated with the experiment can be reformulated in terms of this space. It may seem natural to ask for the numerical value of the probability $\mathbb{P}(A)$ of some event A. The answer to such a question must be contained in the description of the experiment in question. For example, the assertion that a *fair* coin is tossed once is equivalent to saying that heads and tails have an equal probability of occurring; actually, this is the definition of fairness.

(2) Example. A coin, possibly biased, is tossed once. We can take $\Omega = \{\mathrm{H}, \mathrm{T}\}$ and $\mathcal{F} = \{\varnothing, \mathrm{H}, \mathrm{T}, \Omega\}$, and a possible probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ is given by

$$\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\mathrm{H}) = p, \quad \mathbb{P}(\mathrm{T}) = 1 - p, \quad \mathbb{P}(\Omega) = 1,$$

where p is a fixed real number in the interval $[0, 1]$. If $p = \frac{1}{2}$, then we say that the coin is *fair*, or *unbiased*. ●

(3) Example. A die is thrown once. We can take $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \{0, 1\}^{\Omega}$, and the probability measure $\mathbb{P}$ given by

$$\mathbb{P}(A) = \sum_{i \in A} p_i \quad \text{for any } A \subseteq \Omega,$$

where $p_1, p_2, \dots, p_6$ are specified numbers from the interval $[0, 1]$ having unit sum. The probability that i turns up is p_i. The die is fair if $p_i = \frac{1}{6}$ for each i, in which case

$$\mathbb{P}(A) = \tfrac{1}{6}|A| \quad \text{for any } A \subseteq \Omega,$$

where $|A|$ denotes the cardinality of A. ●

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a typical probability space. We now give some of its simple but important properties.

(4) Lemma.

(a) $\mathbb{P}(A^{\mathrm{c}}) = 1 - \mathbb{P}(A)$,

(b) *if* $B \supseteq A$ *then* $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$,

(c) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

(d) *more generally, if* $A_1, A_2, \dots, A_n$ *are events, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$

where, for example, $\sum_{i<j}$ *sums over all unordered pairs* (i, j) *with* $i \neq j$.

Proof.

(a) $A \cup A^{\mathrm{c}} = \Omega$ and $A \cap A^{\mathrm{c}} = \varnothing$, so $\mathbb{P}(A \cup A^{\mathrm{c}}) = \mathbb{P}(A) + \mathbb{P}(A^{\mathrm{c}}) = 1$.

(b) $B = A \cup (B \setminus A)$. This is the union of disjoint sets and therefore

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

(c) $A \cup B = A \cup (B \setminus A)$, which is a disjoint union. Therefore, by (b),

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(d) The proof is by induction on n, and is left as an *exercise* (see Exercise (1.3.4)). ■

In Lemma (4b), $B \setminus A$ denotes the set of members of B which are not in A. In order to write down the quantity $\mathbb{P}(B \setminus A)$, we require that $B \setminus A$ belongs to $\mathcal{F}$, the domain of $\mathbb{P}$; this is always true when A and B belong to $\mathcal{F}$, and to prove this was part of Exercise (1.2.2). Notice that each proof proceeded by expressing an event in terms of disjoint unions and then applying $\mathbb{P}$. It is sometimes easier to calculate the probabilities of intersections of events rather than their unions; part (d) of the lemma is useful then, as we shall discover soon. The next property of $\mathbb{P}$ is more technical, and says that $\mathbb{P}$ is a *continuous* set function; this property is essentially equivalent to the condition that $\mathbb{P}$ is countably additive rather than just finitely additive (see Problem (1.8.16) also).

(5) Lemma. *Let $A_1, A_2, \dots$ be an increasing sequence of events, so that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, and write A for their limit:*

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i.$$

Then $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$.

Similarly, if $B_1, B_2, \dots$ is a decreasing sequence of events, so that $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$, then

$$B = \bigcap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i$$

satisfies $\mathbb{P}(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$.

Proof. $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$ is the union of a disjoint family of events. Thus, by Definition (1),

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A_1) + \sum_{i=1}^{\infty} \mathbb{P}(A_{i+1} \setminus A_i) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n-1} \big[\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)\big] \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}$$

To show the result for decreasing families of events, take complements and use the first part (*exercise*). ■

To recapitulate, statements concerning chance are implicitly related to experiments or trials, the outcomes of which are not entirely predictable. With any such experiment we can associate a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ the properties of which are consistent with our shared and reasonable conceptions of the notion of chance.

Here is some final jargon. An event A is called *null* if $\mathbb{P}(A) = 0$. If $\mathbb{P}(A) = 1$, we say that A occurs *almost surely*. Null events should not be confused with the impossible event $\varnothing$. Null events are happening all around us, even though they have zero probability; after all, what is the chance that a dart strikes any given point of the target at which it is thrown? That is, the impossible event is null, but null events need not be impossible.

Exercises for Section 1.3

1. Let A and B be events with probabilities $\mathbb{P}(A) = \frac{3}{4}$ and $\mathbb{P}(B) = \frac{1}{3}$. Show that $\frac{1}{12} \le \mathbb{P}(A \cap B) \le \frac{1}{3}$, and give examples to show that both extremes are possible. Find corresponding bounds for $\mathbb{P}(A \cup B)$.

2. A fair coin is tossed repeatedly. Show that, with probability one, a head turns up sooner or later. Show similarly that any given finite sequence of heads and tails occurs eventually with probability one. Explain the connection with Murphy's Law.

3. Six cups and saucers come in pairs: there are two cups and saucers which are red, two white, and two with stars on. If the cups are placed randomly onto the saucers (one each), find the probability that no cup is upon a saucer of the same pattern.

4. Let $A_1, A_2, \dots, A_n$ be events where $n \ge 2$, and prove that

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k)$$
$$- \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).$$

In each packet of Corn Flakes may be found a plastic bust of one of the last five Vice-Chancellors of Cambridge University, the probability that any given packet contains any specific Vice-Chancellor being $\frac{1}{5}$, independently of all other packets. Show that the probability that each of the last three Vice-Chancellors is obtained in a bulk purchase of six packets is $1 - 3(\frac{4}{5})^6 + 3(\frac{3}{5})^6 - (\frac{2}{5})^6$.

5. Let $A_r, r \ge 1$, be events such that $\mathbb{P}(A_r) = 1$ for all r. Show that $\mathbb{P}\left(\bigcap_{r=1}^{\infty} A_r\right) = 1$.

6. You are given that at least one of the events A_r, $1 \le r \le n$, is certain to occur, but certainly no more than two occur. If $\mathbb{P}(A_r) = p$, and $\mathbb{P}(A_r \cap A_s) = q$, $r \ne s$, show that $p \ge 1/n$ and $q \le 2/n$.

7. You are given that at least one, but no more than three, of the events A_r, $1 \le r \le n$, occur, where $n \ge 3$. The probability of at least two occurring is $\frac{1}{2}$. If $\mathbb{P}(A_r) = p$, $\mathbb{P}(A_r \cap A_s) = q$, $r \ne s$, and $\mathbb{P}(A_r \cap A_s \cap A_t) = x$, $r < s < t$, show that $p \ge 3/(2n)$, and $q \le 4/n$.

1.4 Conditional probability

Many statements about chance take the form 'if B occurs, then the probability of A is p', where B and A are events (such as 'it rains tomorrow' and 'the bus being on time' respectively) and p is a likelihood as before. To include this in our theory, we return briefly to the discussion about proportions at the beginning of the previous section. An experiment is repeated N times, and on each occasion we observe the occurrences or non-occurrences of two events A and B. Now, suppose we only take an interest in those outcomes for which B occurs; all other experiments are disregarded. In this smaller collection of trials the proportion of times that A occurs is $N(A \cap B)/N(B)$, since B occurs at each of them. However,

$$\frac{N(A \cap B)}{N(B)} = \frac{N(A \cap B)/N}{N(B)/N}.$$

If we now think of these ratios as probabilities, we see that the probability that A occurs, given that B occurs, should be reasonably defined as $\mathbb{P}(A \cap B)/\mathbb{P}(B)$.

Probabilistic intuition leads to the same conclusion. Given that an event B occurs, it is the case that A occurs if and only if $A \cap B$ occurs. Thus the conditional probability of A given B

should be proportional to $\mathbb{P}(A \cap B)$, which is to say that it equals $\alpha\mathbb{P}(A \cap B)$ for some constant $\alpha = \alpha(B)$. The conditional probability of Ω given B must equal 1, and thus $\alpha\mathbb{P}(\Omega \cap B) = 1$, yielding $\alpha = 1/\mathbb{P}(B)$.

We formalize these notions as follows.

(1) Definition. If $\mathbb{P}(B) > 0$ then the **conditional probability** that A occurs given that B occurs is defined to be

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

We denote this conditional probability by $\mathbb{P}(A \mid B)$, pronounced 'the probability of A given B', or sometimes 'the probability of A conditioned (or conditional) on B'.

(2) Example. Two fair dice are thrown. Given that the first shows 3, what is the probability that the total exceeds 6? The answer is obviously $\frac{1}{2}$, since the second must show 4, 5, or 6. However, let us labour the point. Clearly $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, the set† of all ordered pairs (i, j) for $i, j \in \{1, 2, \dots, 6\}$, and we can take $\mathcal{F}$ to be the set of all subsets of Ω, with $\mathbb{P}(A) = |A|/36$ for any $A \subseteq \Omega$. Let B be the event that the first die shows 3, and A be the event that the total exceeds 6. Then

$$B = \{(3, b) : 1 \le b \le 6\}, \quad A = \{(a, b) : a + b > 6\}, \quad A \cap B = \{(3, 4), (3, 5), (3, 6)\},$$

and

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{|A \cap B|}{|B|} = \frac{3}{6}. \qquad \bullet$$

(3) Example. A family has two children. What is the probability that both are boys, given that at least one is a boy? The older and younger child may each be male or female, so there are four possible combinations of sexes, which we assume to be equally likely. Hence we can represent the sample space in the obvious way as

$$\Omega = \{\text{GG}, \text{GB}, \text{BG}, \text{BB}\}$$

where $\mathbb{P}(\text{GG}) = \mathbb{P}(\text{BB}) = \mathbb{P}(\text{GB}) = \mathbb{P}(\text{BG}) = \frac{1}{4}$. From the definition of conditional probability,

$$\begin{aligned}\mathbb{P}(\text{BB} \mid \text{one boy at least}) &= \mathbb{P}(\text{BB} \mid \text{GB} \cup \text{BG} \cup \text{BB}) \\ &= \frac{\mathbb{P}(\text{BB} \cap (\text{GB} \cup \text{BG} \cup \text{BB}))}{\mathbb{P}(\text{GB} \cup \text{BG} \cup \text{BB})} \\ &= \frac{\mathbb{P}(\text{BB})}{\mathbb{P}(\text{GB} \cup \text{BG} \cup \text{BB})} = \frac{1}{3}.\end{aligned}$$

A popular but incorrect answer to the question is $\frac{1}{2}$. This is the correct answer to another question: for a family with two children, what is the probability that both are boys given that the younger is a boy? In this case,

$$\begin{aligned}\mathbb{P}(\text{BB} \mid \text{younger is a boy}) &= \mathbb{P}(\text{BB} \mid \text{GB} \cup \text{BB}) \\ &= \frac{\mathbb{P}(\text{BB} \cap (\text{GB} \cup \text{BB}))}{\mathbb{P}(\text{GB} \cup \text{BB})} = \frac{\mathbb{P}(\text{BB})}{\mathbb{P}(\text{GB} \cup \text{BB})} = \frac{1}{2}.\end{aligned}$$

†Remember that $A \times B = \{(a, b) : a \in A,\ b \in B\}$ and that $A \times A = A^2$.

The usual dangerous argument contains the assertion

$$\mathbb{P}(\text{BB} \mid \text{one child is a boy}) = \mathbb{P}(\text{other child is a boy}).$$

Why is this meaningless? [Hint: Consider the sample space.] ●

The next lemma is crucially important in probability theory. A family $B_1, B_2, \dots, B_n$ of events is called a *partition* of the set Ω if

$$B_i \cap B_j = \varnothing \quad \text{when} \quad i \neq j, \quad \text{and} \quad \bigcup_{i=1}^{n} B_i = \Omega.$$

Each elementary event $\omega \in \Omega$ belongs to exactly one set in a partition of Ω.

(4) Lemma. *For any events A and B such that $0 < \mathbb{P}(B) < 1$,*

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^{c})\mathbb{P}(B^{c}).$$

More generally, let $B_1, B_2, \dots, B_n$ be a partition of Ω such that $\mathbb{P}(B_i) > 0$ for all i. Then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Proof. $A = (A \cap B) \cup (A \cap B^{c})$. This is a disjoint union and so

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^{c}) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^{c})\mathbb{P}(B^{c}).\end{aligned}$$

The second part is similar (see Problem (1.8.10)). ■

(5) Example. We are given two urns, each containing a collection of coloured balls. Urn I contains two white and three blue balls, whilst urn II contains three white and four blue balls. A ball is drawn at random from urn I and put into urn II, and then a ball is picked at random from urn II and examined. What is the probability that it is blue? We assume unless otherwise specified that a ball picked randomly from any urn is equally likely to be any of those present. The reader will be relieved to know that we no longer need to describe $(\Omega, \mathcal{F}, \mathbb{P})$ in detail; we are confident that we could do so if necessary. Clearly, the colour of the final ball depends on the colour of the ball picked from urn I. So let us 'condition' on this. Let A be the event that the final ball is blue, and let B be the event that the first one picked was blue. Then, by Lemma (4),

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^{c})\mathbb{P}(B^{c}).$$

We can easily find all these probabilities:

$$\begin{aligned}\mathbb{P}(A \mid B) &= \mathbb{P}(A \mid \text{urn II contains three white and five blue balls}) = \tfrac{5}{8}, \\ \mathbb{P}(A \mid B^{c}) &= \mathbb{P}(A \mid \text{urn II contains four white and four blue balls}) = \tfrac{1}{2}, \\ \mathbb{P}(B) &= \tfrac{3}{5}, \qquad \mathbb{P}(B^{c}) = \tfrac{2}{5}.\end{aligned}$$

Hence

$$\mathbb{P}(A) = \tfrac{5}{8} \cdot \tfrac{3}{5} + \tfrac{1}{2} \cdot \tfrac{2}{5} = \tfrac{23}{40}.$$ ●

Unprepared readers may have been surprised by the sudden appearance of urns in this book. In the seventeenth and eighteenth centuries, lotteries often involved the drawing of slips from urns, and voting was often a matter of putting slips or balls into urns. In France today, *aller aux urnes* is synonymous with voting. It was therefore not unnatural for the numerous Bernoullis and others to model births, marriages, deaths, fluids, gases, and so on, using urns containing balls of varied hue.

(6) Example. Only two factories manufacture zoggles. 20 per cent of the zoggles from factory I and 5 per cent from factory II are defective. Factory I produces twice as many zoggles as factory II each week. What is the probability that a zoggle, randomly chosen from a week's production, is satisfactory? Clearly this satisfaction depends on the factory of origin. Let A be the event that the chosen zoggle is satisfactory, and let B be the event that it was made in factory I. Arguing as before,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^{c})\mathbb{P}(B^{c}) \\ &= \tfrac{4}{5} \cdot \tfrac{2}{3} + \tfrac{19}{20} \cdot \tfrac{1}{3} = \tfrac{51}{60}.\end{aligned}$$

If the chosen zoggle is defective, what is the probability that it came from factory I? In our notation this is just $\mathbb{P}(B \mid A^{c})$. However,

$$\mathbb{P}(B \mid A^{c}) = \frac{\mathbb{P}(B \cap A^{c})}{\mathbb{P}(A^{c})} = \frac{\mathbb{P}(A^{c} \mid B)\mathbb{P}(B)}{\mathbb{P}(A^{c})} = \frac{\frac{1}{5} \cdot \frac{2}{3}}{1 - \frac{51}{60}} = \frac{8}{9}.$$ ●

This section is terminated with a cautionary example. It is not untraditional to perpetuate errors of logic in calculating conditional probabilities. Lack of unambiguous definitions and notation has led astray many probabilists, including even Boole, who was credited by Russell with the discovery of pure mathematics and by others for some of the logical foundations of computing. The well-known 'prisoners' paradox' also illustrates some of the dangers here.

(7) Example. Prisoners' paradox. In a dark country, three prisoners have been incarcerated without trial. Their warder tells them that the country's dictator has decided arbitrarily to free one of them and to shoot the other two, but he is not permitted to reveal to any prisoner the fate of that prisoner. Prisoner A knows therefore that his chance of survival is $\frac{1}{3}$. In order to gain information, he asks the warder to tell him in secret the name of some prisoner (but not himself) who will be killed, and the warder names prisoner B. What now is prisoner A's assessment of the chance that he will survive? Could it be $\frac{1}{2}$: after all, he knows now that the survivor will be either A or C, and he has no information about which? Could it be $\frac{1}{3}$: after all, according to the rules, at least one of B and C has to be killed, and thus the extra information cannot reasonably affect A's earlier calculation of the odds? What does the reader think about this? The resolution of the paradox lies in the situation when either response (B or C) is possible.

An alternative formulation of this paradox has become known as the Monty Hall problem, the controversy associated with which has been provoked by Marilyn vos Savant (and many others) in *Parade* magazine in 1990; see Exercise (1.4.5). ●

Exercises for Section 1.4

1. Prove that $\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\mathbb{P}(A)/\mathbb{P}(B)$ whenever $\mathbb{P}(A)\mathbb{P}(B) \neq 0$. Show that, if $\mathbb{P}(A \mid B) > \mathbb{P}(A)$, then $\mathbb{P}(B \mid A) > \mathbb{P}(B)$.

2. For events $A_1, A_2, \dots, A_n$ satisfying $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, prove that

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

3. A man possesses five coins, two of which are double-headed, one is double-tailed, and two are normal. He shuts his eyes, picks a coin at random, and tosses it. What is the probability that the lower face of the coin is a head?

He opens his eyes and sees that the coin is showing heads; what is the probability that the lower face is a head?

He shuts his eyes again, and tosses the coin again. What is the probability that the lower face is a head?

He opens his eyes and sees that the coin is showing heads; what is the probability that the lower face is a head?

He discards this coin, picks another at random, and tosses it. What is the probability that it shows heads?

4. What do you think of the following 'proof' by Lewis Carroll that an urn cannot contain two balls of the same colour? Suppose that the urn contains two balls, each of which is either black or white; thus, in the obvious notation, $\mathbb{P}(\mathrm{BB}) = \mathbb{P}(\mathrm{BW}) = \mathbb{P}(\mathrm{WB}) = \mathbb{P}(\mathrm{WW}) = \frac{1}{4}$. We add a black ball, so that $\mathbb{P}(\mathrm{BBB}) = \mathbb{P}(\mathrm{BBW}) = \mathbb{P}(\mathrm{BWB}) = \mathbb{P}(\mathrm{BWW}) = \frac{1}{4}$. Next we pick a ball at random; the chance that the ball is black is (using conditional probabilities) $1 \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} + \frac{1}{3} \cdot \frac{1}{4} = \frac{2}{3}$. However, if there is probability $\frac{2}{3}$ that a ball, chosen randomly from three, is black, then there must be two black and one white, which is to say that originally there was one black and one white ball in the urn.

5. **The Monty Hall problem: goats and cars.** (a) Cruel fate has made you a contestant in a game show; you have to choose one of three doors. One conceals a new car, two conceal old goats. You choose, but your chosen door is not opened immediately. Instead, the presenter opens another door to reveal a goat, and he offers you the opportunity to change your choice to the third door (unopened and so far unchosen). Let p be the (conditional) probability that the third door conceals the car. The value of p depends on the presenter's protocol. Devise protocols to yield the values $p = \frac{1}{2}$, $p = \frac{2}{3}$. Show that, for $\alpha \in [\frac{1}{2}, \frac{2}{3}]$, there exists a protocol such that $p = \alpha$. Are you well advised to change your choice to the third door?

(b) In a variant of this question, the presenter is permitted to open the first door chosen, and to reward you with whatever lies behind. If he chooses to open another door, then this door invariably conceals a goat. Let p be the probability that the unopened door conceals the car, conditional on the presenter having chosen to open a second door. Devise protocols to yield the values $p = 0$, $p = 1$, and deduce that, for any $\alpha \in [0, 1]$, there exists a protocol with $p = \alpha$.

6. **The prosecutor's fallacy†.** Let G be the event that an accused is guilty, and T the event that some testimony is true. Some lawyers have argued on the assumption that $\mathbb{P}(G \mid T) = \mathbb{P}(T \mid G)$. Show that this holds if and only if $\mathbb{P}(G) = \mathbb{P}(T)$.

7. **Urns.** There are n urns of which the rth contains $r - 1$ red balls and $n - r$ magenta balls. You pick an urn at random and remove two balls at random without replacement. Find the probability that:

(a) the second ball is magenta;

(b) the second ball is magenta, given that the first is magenta.

†The prosecution made this error in the famous Dreyfus case of 1894.

1.5 Independence

In general, the occurrence of some event B changes the probability that another event A occurs, the original probability $\mathbb{P}(A)$ being replaced by $\mathbb{P}(A \mid B)$. If the probability remains unchanged, that is to say $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, then we call A and B 'independent'. This is well defined only if $\mathbb{P}(B) > 0$. Definition (1.4.1) of conditional probability leads us to the following.

(1) Definition. Events A and B are called **independent** if

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

More generally, a family $\{A_i : i \in I\}$ is called **independent** if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

for all finite subsets J of I.

Remark. A common student error is to make the fallacious statement that A and B are independent if $A \cap B = \varnothing$.

If the family $\{A_i : i \in I\}$ has the property that

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \qquad \text{for all } i \neq j$$

then it is called *pairwise independent*. Pairwise-independent families are not necessarily independent, as the following example shows.

(2) Example. Suppose $\Omega = \{abc, acb, cab, cba, bca, bac, aaa, bbb, ccc\}$, and each of the nine elementary events in Ω occurs with equal probability $\frac{1}{9}$. Let A_k be the event that the kth letter is a. It is left as an *exercise* to show that the family $\{A_1, A_2, A_3\}$ is pairwise independent but not independent. ●

(3) Example (1.4.6) revisited. The events A and B of this example are clearly dependent because $\mathbb{P}(A \mid B) = \frac{4}{5}$ and $\mathbb{P}(A) = \frac{51}{60}$. ●

(4) Example. Choose a card at random from a pack of 52 playing cards, each being picked with equal probability $\frac{1}{52}$. We claim that the suit of the chosen card is independent of its rank. For example,

$$\mathbb{P}(\text{king}) = \tfrac{4}{52}, \qquad \mathbb{P}(\text{king} \mid \text{spade}) = \tfrac{1}{13}.$$

Alternatively,

$$\mathbb{P}(\text{spade king}) = \tfrac{1}{52} = \tfrac{1}{4} \cdot \tfrac{1}{13} = \mathbb{P}(\text{spade})\mathbb{P}(\text{king}).$$ ●

Let C be an event with $\mathbb{P}(C) > 0$. To the conditional probability measure $\mathbb{P}(\cdot \mid C)$ corresponds the idea of *conditional independence*. Two events A and B are called *conditionally independent given* C if

(5)
$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C);$$

there is a natural extension to families of events. [However, note Exercise (1.5.5).]

Exercises for Section 1.5

1. Let A and B be independent events; show that A^c, B are independent, and deduce that A^c, B^c are independent.

2. We roll a die n times. Let A_{ij} be the event that the ith and jth rolls produce the same number. Show that the events $\{A_{ij} : 1 \le i < j \le n\}$ are pairwise independent but not independent.

3. A fair coin is tossed repeatedly. Show that the following two statements are equivalent:
(a) the outcomes of different tosses are independent,
(b) for any given finite sequence of heads and tails, the chance of this sequence occurring in the first m tosses is 2^{-m}, where m is the length of the sequence.

4. Let $\Omega = \{1, 2, \dots, p\}$ where p is prime, $\mathcal{F}$ be the set of all subsets of Ω, and $\mathbb{P}(A) = |A|/p$ for all $A \in \mathcal{F}$. Show that, if A and B are independent events, then at least one of A and B is either $\varnothing$ or Ω.

5. Show that the conditional independence of A and B given C neither implies, nor is implied by, the independence of A and B. For which events C is it the case that, for all A and B, the events A and B are independent if and only if they are conditionally independent given C?

6. **Safe or sorry?** Some form of prophylaxis is said to be 90 per cent effective at prevention during one year's treatment. If the degrees of effectiveness in different years are independent, show that the treatment is more likely than not to fail within 7 years.

7. **Families.** Jane has three children, each of which is equally likely to be a boy or a girl independently of the others. Define the events:

$$\begin{aligned} A &= \{\text{all the children are of the same sex}\}, \\ B &= \{\text{there is at most one boy}\}, \\ C &= \{\text{the family includes a boy and a girl}\}. \end{aligned}$$

(a) Show that A is independent of B, and that B is independent of C.
(b) Is A independent of C?
(c) Do these results hold if boys and girls are not equally likely?
(d) Do these results hold if Jane has four children?

8. **Galton's paradox.** You flip three fair coins. At least two are alike, and it is an evens chance that the third is a head or a tail. Therefore $\mathbb{P}(\text{all alike}) = \frac{1}{2}$. Do you agree?

9. Two fair dice are rolled. Show that the event that their sum is 7 is independent of the score shown by the first die.

1.6 Completeness and product spaces

This section should be omitted at the first reading, but we shall require its contents later. It contains only a sketch of complete probability spaces and product spaces; the reader should look elsewhere for a more detailed treatment (see Billingsley 1995). We require the following result.

(1) Lemma. *If $\mathcal{F}$ and $\mathcal{G}$ are two σ-fields of subsets of Ω then their intersection $\mathcal{F} \cap \mathcal{G}$ is a σ-field also. More generally, if $\{\mathcal{F}_i : i \in I\}$ is a family of σ-fields of subsets of Ω then $\mathcal{G} = \bigcap_{i \in I} \mathcal{F}_i$ is a σ-field also.*

The proof is not difficult and is left as an *exercise*. Note that the union $\mathcal{F} \cup \mathcal{G}$ may not be a σ-field, although it may be extended to a unique smallest σ-field written $\sigma(\mathcal{F} \cup \mathcal{G})$, as follows. Let $\{\mathcal{G}_i : i \in I\}$ be the collection of all σ-fields which contain both $\mathcal{F}$ and $\mathcal{G}$ as subsets; this collection is non-empty since it contains the set of all subsets of Ω. Then $\mathcal{G} = \bigcap_{i \in I} \mathcal{G}_i$ is the unique smallest σ-field which contains $\mathcal{F} \cup \mathcal{G}$.

(A) Completeness. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Any event A which has zero probability, that is $\mathbb{P}(A) = 0$, is called *null*. It may seem reasonable to suppose that any subset B of a null set A will itself be null, but this may be without meaning since B may not be an event, and thus $\mathbb{P}(B)$ may not be defined.

(2) Definition. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called **complete** if all subsets of null sets are events.

Any incomplete space can be completed thus. Let $\mathcal{N}$ be the collection of all subsets of null sets in $\mathcal{F}$ and let $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N})$ be the smallest σ-field which contains all sets in $\mathcal{F}$ and $\mathcal{N}$. It can be shown that the domain of $\mathbb{P}$ may be extended in an obvious way from $\mathcal{F}$ to $\mathcal{G}$; $(\Omega, \mathcal{G}, \mathbb{P})$ is called the *completion* of $(\Omega, \mathcal{F}, \mathbb{P})$.

(B) Product spaces. The probability spaces discussed in this chapter have usually been constructed around the outcomes of one experiment, but instances occur naturally when we need to combine the outcomes of several independent experiments into one space (see Examples (1.2.4) and (1.4.2)). How should we proceed in general?

Suppose two experiments have associated probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ respectively. The sample space of the pair of experiments, considered jointly, is the collection $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$ of ordered pairs. The appropriate σ-field of events is more complicated to construct. Certainly it should contain all subsets of $\Omega_1 \times \Omega_2$ of the form $A_1 \times A_2 = \{(a_1, a_2) : a_1 \in A_1,\ a_2 \in A_2\}$ where A_1 and A_2 are typical members of $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. However, the family of all such sets, $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}$, is not in general a σ-field. By the discussion after (1), there exists a unique smallest σ-field $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ of subsets of $\Omega_1 \times \Omega_2$ which contains $\mathcal{F}_1 \times \mathcal{F}_2$. All we require now is a suitable probability function on $(\Omega_1 \times \Omega_2, \mathcal{G})$. Let $\mathbb{P}_{12} : \mathcal{F}_1 \times \mathcal{F}_2 \to [0, 1]$ be given by:

(3) $$\mathbb{P}_{12}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \quad \text{for } A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2.$$

It can be shown that the domain of $\mathbb{P}_{12}$ can be extended from $\mathcal{F}_1 \times \mathcal{F}_2$ to the whole of $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. The ensuing probability space $(\Omega_1 \times \Omega_2, \mathcal{G}, \mathbb{P}_{12})$ is called the *product space* of $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Products of larger numbers of spaces are constructed similarly. The measure $\mathbb{P}_{12}$ is sometimes called the 'product measure' since its defining equation (3) assumed that two experiments are independent. There are of course many other measures that can be applied to $(\Omega_1 \times \Omega_2, \mathcal{G})$.

In many simple cases this technical discussion is unnecessary. Suppose that Ω_1 and Ω_2 are finite, and that their σ-fields contain all their subsets; this is the case in Examples (1.2.4) and (1.4.2). Then $\mathcal{G}$ contains all subsets of $\Omega_1 \times \Omega_2$.

1.7 Worked examples

Here are some more examples to illustrate the ideas of this chapter. The reader is now equipped to try his or her hand at a substantial number of those problems which exercised the pioneers in probability. These frequently involved experiments having equally likely outcomes, such as dealing whist hands, putting balls of various colours into urns and taking them out again, throwing dice, and so on. In many such instances, the reader will be pleasantly surprised to find that it is not necessary to write down $(\Omega, \mathcal{F}, \mathbb{P})$ explicitly, but only to think of Ω as being a collection $\{\omega_1, \omega_2, \dots, \omega_N\}$ of possibilities, each of which may occur with probability $1/N$. Thus, $\mathbb{P}(A) = |A|/N$ for any $A \subseteq \Omega$. The basic tools used in such problems are as follows.

(a) Combinatorics: remember that the number of permutations of n objects is $n!$ and that the number of ways of choosing r objects from n is $\binom{n}{r}$.

(b) Set theory: to obtain $\mathbb{P}(A)$ we can compute $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ or we can partition A by conditioning on events B_i, and then use Lemma (1.4.4).

(c) Use of independence.

(1) Example. Consider a series of hands dealt at bridge. Let A be the event that in a given deal each player has one ace. Show that the probability that A occurs at least once in seven deals is approximately $\frac{1}{2}$.

Solution. The number of ways of dealing 52 cards into four equal hands is $52!/(13!)^4$. There are $4!$ ways of distributing the aces so that each hand holds one, and there are $48!/(12!)^4$ ways of dealing the remaining cards. Thus

$$\mathbb{P}(A) = \frac{4!\,48!/(12!)^4}{52!/(13!)^4} \simeq \frac{1}{10}.$$

Now let B_i be the event that A occurs for the first time on the ith deal. Clearly $B_i \cap B_j = \varnothing$, $i \neq j$. Thus

$$\mathbb{P}(A \text{ occurs in seven deals}) = \mathbb{P}(B_1 \cup \dots \cup B_7) = \sum_1^7 \mathbb{P}(B_i) \quad \text{using Definition (1.3.1).}$$

Since successive deals are independent, we have

$$\begin{aligned}\mathbb{P}(B_i) &= \mathbb{P}\big(A^c \text{ occurs on deal } 1,\ A^c \text{ occurs on deal } 2, \\ &\qquad \dots,\ A^c \text{ occurs on deal } i-1,\ \ A \text{ occurs on deal } i\big) \\ &= \mathbb{P}(A^c)^{i-1}\mathbb{P}(A) \quad \text{using Definition (1.5.1)} \\ &\simeq \left(1 - \tfrac{1}{10}\right)^{i-1} \tfrac{1}{10}.\end{aligned}$$

Thus

$$\mathbb{P}(A \text{ occurs in seven deals}) = \sum_1^7 \mathbb{P}(B_i) \simeq \sum_1^7 \left(\tfrac{9}{10}\right)^{i-1} \tfrac{1}{10} \simeq \tfrac{1}{2}.$$

Can you see an easier way of obtaining this answer? ●

(2) Example. There are two roads from A to B and two roads from B to C. Each of the four roads has probability p of being blocked by snow, independently of all the others. What is the probability that there is an open road from A to C?

Solution.

$$\begin{aligned}\mathbb{P}(\text{open road}) &= \mathbb{P}\big((\text{open road from A to B}) \cap (\text{open road from B to C})\big)\\ &= \mathbb{P}(\text{open road from A to B})\mathbb{P}(\text{open road from B to C})\end{aligned}$$

using the independence. However, p is the same for all roads; thus, using Lemma (1.3.4),

$$\begin{aligned}\mathbb{P}(\text{open road}) &= \big(1 - \mathbb{P}(\text{no road from A to B})\big)^2\\ &= \big\{1 - \mathbb{P}\big((\text{first road blocked}) \cap (\text{second road blocked})\big)\big\}^2\\ &= \big\{1 - \mathbb{P}(\text{first road blocked})\mathbb{P}(\text{second road blocked})\big\}^2\end{aligned}$$

using the independence. Thus

(3) $$\mathbb{P}(\text{open road}) = (1 - p^2)^2.$$

Further suppose that there is also a direct road from A to C, which is independently blocked with probability p. Then, by Lemma (1.4.4) and equation (3),

$$\begin{aligned}\mathbb{P}(\text{open road}) &= \mathbb{P}(\text{open road} \mid \text{direct road blocked}) \cdot p\\ &\qquad + \mathbb{P}(\text{open road} \mid \text{direct road open}) \cdot (1 - p)\\ &= (1 - p^2)^2 \cdot p + 1 \cdot (1 - p).\end{aligned}$$ ●

(4) Example. Symmetric random walk (or 'Gambler's ruin'). A man is saving up to buy a new Jaguar at a cost of N units of money. He starts with k units where $0 < k < N$, and tries to win the remainder by the following gamble with his bank manager. He tosses a fair coin repeatedly; if it comes up heads then the manager pays him one unit, but if it comes up tails then he pays the manager one unit. He plays this game repeatedly until one of two events occurs: either he runs out of money and is bankrupted or he wins enough to buy the Jaguar. What is the probability that he is ultimately bankrupted?

Solution. This is one of many problems the solution to which proceeds by the construction of a linear difference equation subject to certain boundary conditions. Let A denote the event that he is eventually bankrupted, and let B be the event that the first toss of the coin shows heads. By Lemma (1.4.4),

(5) $$\mathbb{P}_k(A) = \mathbb{P}_k(A \mid B)\mathbb{P}(B) + \mathbb{P}_k(A \mid B^{\mathrm{c}})\mathbb{P}(B^{\mathrm{c}}),$$

where $\mathbb{P}_k$ denotes probabilities calculated relative to the starting point k. We want to find $\mathbb{P}_k(A)$. Consider $\mathbb{P}_k(A \mid B)$. If the first toss is a head then his capital increases to $k + 1$ units and the game starts afresh from a different starting point. Thus $\mathbb{P}_k(A \mid B) = \mathbb{P}_{k+1}(A)$ and similarly $\mathbb{P}_k(A \mid B^{\mathrm{c}}) = \mathbb{P}_{k-1}(A)$. So, writing $p_k = \mathbb{P}_k(A)$, (5) becomes

(6) $$p_k = \tfrac{1}{2}(p_{k+1} + p_{k-1}) \quad \text{if} \quad 0 < k < N,$$

which is a linear difference equation subject to the boundary conditions $p_0 = 1$, $p_N = 0$. The analytical solution to such equations is routine, and we shall return later to the general method of solution. In this case we can proceed directly. We put $b_k = p_k - p_{k-1}$ to obtain $b_k = b_{k-1}$ and hence $b_k = b_1$ for all k. Thus

$$p_k = b_1 + p_{k-1} = 2b_1 + p_{k-2} = \cdots = kb_1 + p_0$$

is the general solution to (6). The boundary conditions imply that $p_0 = 1$, $b_1 = -1/N$, giving

$$\mathbb{P}_k(A) = 1 - \frac{k}{N}. \tag{7}$$

As the price of the Jaguar rises, that is as $N \to \infty$, ultimate bankruptcy becomes very likely. This is the problem of the 'symmetric random walk with two absorbing barriers' to which we shall return in more generality later. ●

Remark. Our experience of student calculations leads us to stress that probabilities lie between zero and one; any calculated probability which violates this must be incorrect.

(8) Example. Testimony. A court is investigating the possible occurrence of an unlikely event T. The reliability of two independent witnesses called Alf and Bob is known to the court: Alf tells the truth with probability α and Bob with probability β, and there is no collusion between the two of them. Let A and B be the events that Alf and Bob assert (respectively) that T occurred, and let $\tau = \mathbb{P}(T)$. What is the probability that T occurred given that both Alf and Bob declare that T occurred?

Solution. We are asked to calculate $\mathbb{P}(T \mid A \cap B)$, which is equal to $\mathbb{P}(T \cap A \cap B)/\mathbb{P}(A \cap B)$. Now $\mathbb{P}(T \cap A \cap B) = \mathbb{P}(A \cap B \mid T)\mathbb{P}(T)$ and

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap B \mid T)\mathbb{P}(T) + \mathbb{P}(A \cap B \mid T^c)\mathbb{P}(T^c).$$

We have from the independence of the witnesses that A and B are conditionally independent given either T or T^c. Therefore

$$\mathbb{P}(A \cap B \mid T) = \mathbb{P}(A \mid T)\mathbb{P}(B \mid T) = \alpha\beta,$$
$$\mathbb{P}(A \cap B \mid T^c) = \mathbb{P}(A \mid T^c)\mathbb{P}(B \mid T^c) = (1-\alpha)(1-\beta),$$

so that

$$\mathbb{P}(T \mid A \cap B) = \frac{\alpha\beta\tau}{\alpha\beta\tau + (1-\alpha)(1-\beta)(1-\tau)}.$$

As an example, suppose that $\alpha = \beta = \frac{9}{10}$ and $\tau = 1/1000$. Then $\mathbb{P}(T \mid A \cap B) = 81/1080$, which is somewhat small as a basis for a judicial conclusion.

This calculation may be informative. However, it is generally accepted that such an application of the axioms of probability is inappropriate to questions of truth and belief. ●

(9) Example. Zoggles revisited. A new process for the production of zoggles is invented, and both factories of Example (1.4.6) install extra production lines using it. The new process is cheaper but produces fewer reliable zoggles, only 75 per cent of items produced in this new way being reliable.

Factory I fails to implement its new production line efficiently, and only 10 per cent of its output is made in this manner. Factory II does better: it produces 20 per cent of its output by the new technology, and now produces twice as many zoggles in all as Factory I.

Is the new process beneficial to the consumer?

Solution. Both factories now produce a higher proportion of unreliable zoggles than before, and so it might seem at first sight that there is an increased proportion of unreliable zoggles on the market.

Let A be the event that a randomly chosen zoggle is satisfactory, B the event that it came from factory I, and C the event that it was made by the new method. Then

$$\begin{aligned}\mathbb{P}(A) &= \tfrac{1}{3}\mathbb{P}(A \mid B) + \tfrac{2}{3}\mathbb{P}(A \mid B^{c})\\ &= \tfrac{1}{3}\left(\tfrac{1}{10}\mathbb{P}(A \mid B \cap C) + \tfrac{9}{10}\mathbb{P}(A \mid B \cap C^{c})\right)\\ &\qquad + \tfrac{2}{3}\left(\tfrac{1}{5}\mathbb{P}(A \mid B^{c} \cap C) + \tfrac{4}{5}\mathbb{P}(A \mid B^{c} \cap C^{c})\right)\\ &= \tfrac{1}{3}\left(\tfrac{1}{10}\cdot\tfrac{3}{4} + \tfrac{9}{10}\cdot\tfrac{4}{5}\right) + \tfrac{2}{3}\left(\tfrac{1}{5}\cdot\tfrac{3}{4} + \tfrac{4}{5}\cdot\tfrac{19}{20}\right) = \tfrac{523}{600} > \tfrac{51}{60},\end{aligned}$$

so that the proportion of satisfactory zoggles has been increased. ●

(10) Example. Simpson's paradox†. A doctor has performed clinical trials to determine the relative efficacies of two drugs, with the following results.

	Women		Men	
	Drug I	Drug II	Drug I	Drug II
Success	200	10	19	1000
Failure	1800	190	1	1000

Which drug is the better? Here are two conflicting responses.

1. Drug I was given to 2020 people, of whom 219 were cured. The success rate was 219/2020, which is much smaller than the corresponding figure, 1010/2200, for drug II. Therefore drug II is better than drug I.
2. Amongst women the success rates of the drugs are 1/10 and 1/20, and amongst men 19/20 and 1/2. Drug I wins in both cases.

This well-known statistical paradox may be reformulated in the following more general way. Given three events A, B, C, it is possible to allocate probabilities such that

$$\textbf{(11)}\qquad \mathbb{P}(A \mid B \cap C) > \mathbb{P}(A \mid B^{c} \cap C) \quad\text{and}\quad \mathbb{P}(A \mid B \cap C^{c}) > \mathbb{P}(A \mid B^{c} \cap C^{c})$$

but

$$\textbf{(12)}\qquad \mathbb{P}(A \mid B) < \mathbb{P}(A \mid B^{c}).$$

†This paradox, named after Simpson (1951), was remarked by Yule in 1903. The nomenclature is an instance of Stigler's law of eponymy: "No law, theorem, or discovery is named after its originator". This law applies to many eponymous statements in this book, including the law itself. As remarked by A. N. Whitehead, "Everything of importance has been said before, by somebody who did not discover it".

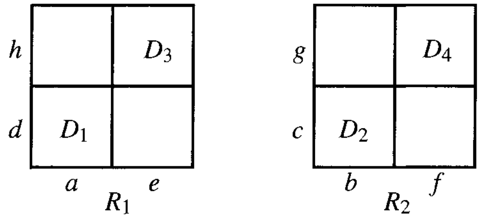

Figure 1.1. Two unions of rectangles illustrating Simpson's paradox.

We may think of A as the event that treatment is successful, B as the event that drug I is given to a randomly chosen individual, and C as the event that this individual is female. The above inequalities imply that B is preferred to B^c when C occurs and when C^c occurs, but B^c is preferred to B overall.

Setting

$$\begin{aligned} a &= \mathbb{P}(A \cap B \cap C), & b &= \mathbb{P}(A^c \cap B \cap C), \\ c &= \mathbb{P}(A \cap B^c \cap C), & d &= \mathbb{P}(A^c \cap B^c \cap C), \\ e &= \mathbb{P}(A \cap B \cap C^c), & f &= \mathbb{P}(A^c \cap B \cap C^c), \\ g &= \mathbb{P}(A \cap B^c \cap C^c), & h &= \mathbb{P}(A^c \cap B^c \cap C^c), \end{aligned}$$

and expanding (11)–(12), we arrive at the (equivalent) inequalities

(13) $$ad > bc, \quad eh > fg, \quad (a+e)(d+h) < (b+f)(c+g),$$

subject to the conditions $a, b, c, \dots, h \geq 0$ and $a+b+c+\cdots+h = 1$. Inequalities (13) are equivalent to the existence of two rectangles R_1 and R_2, as in Figure 1.1, satisfying

$$\text{area}\,(D_1) > \text{area}\,(D_2), \quad \text{area}\,(D_3) > \text{area}\,(D_4), \quad \text{area}\,(R_1) < \text{area}\,(R_2).$$

Many such rectangles may be found, by inspection, as for example those with $a = \frac{3}{30}$, $b = \frac{1}{30}$, $c = \frac{8}{30}$, $d = \frac{3}{30}$, $e = \frac{3}{30}$, $f = \frac{8}{30}$, $g = \frac{1}{30}$, $h = \frac{3}{30}$. Similar conclusions are valid for finer partitions $\{C_i : i \in I\}$ of the sample space, though the corresponding pictures are harder to draw.

Simpson's paradox has arisen many times in practical situations. There are many well-known cases, including the admission of graduate students to the University of California at Berkeley and a clinical trial comparing treatments for kidney stones. ●

(14) Example. False positives. A rare disease affects one person in 10^5. A test for the disease shows positive with probability $\frac{99}{100}$ when applied to an ill person, and with probability $\frac{1}{100}$ when applied to a healthy person. What is the probability that you have the disease given that the test shows positive?

Solution. In the obvious notation,

$$\begin{aligned} \mathbb{P}(\text{ill} \mid +) &= \frac{\mathbb{P}(+ \mid \text{ill})\mathbb{P}(\text{ill})}{\mathbb{P}(+ \mid \text{ill})\mathbb{P}(\text{ill}) + \mathbb{P}(+ \mid \text{healthy})\mathbb{P}(\text{healthy})} \\ &= \frac{\frac{99}{100} \cdot 10^{-5}}{\frac{99}{100} \cdot 10^{-5} + \frac{1}{100}(1 - 10^{-5})} = \frac{99}{99 + 10^5 - 1} \simeq \frac{1}{1011}. \end{aligned}$$

The chance of being ill is rather small. Indeed it is more likely that the test was incorrect. ●

Exercises for Section 1.7

1. There are two roads from A to B and two roads from B to C. Each of the four roads is blocked by snow with probability p, independently of the others. Find the probability that there is an open road from A to B given that there is no open route from A to C.

If, in addition, there is a direct road from A to C, this road being blocked with probability p independently of the others, find the required conditional probability.

2. Calculate the probability that a hand of 13 cards dealt from a normal shuffled pack of 52 contains exactly two kings and one ace. What is the probability that it contains exactly one ace given that it contains exactly two kings?

3. A symmetric random walk takes place on the integers $0, 1, 2, \dots, N$ with absorbing barriers at 0 and N, starting at k. Show that the probability that the walk is never absorbed is zero.

4. The so-called 'sure thing principle' asserts that if you prefer x to y given C, and also prefer x to y given C^{c}, then you surely prefer x to y. Agreed?

5. A pack contains m cards, labelled $1, 2, \dots, m$. The cards are dealt out in a random order, one by one. Given that the label of the kth card dealt is the largest of the first k cards dealt, what is the probability that it is also the largest in the pack?

1.8 Problems

1. A traditional fair die is thrown twice. What is the probability that:
(a) a six turns up exactly once?
(b) both numbers are odd?
(c) the sum of the scores is 4?
(d) the sum of the scores is divisible by 3?

2. A fair coin is thrown repeatedly. What is the probability that on the nth throw:
(a) a head appears for the first time?
(b) the numbers of heads and tails to date are equal?
(c) exactly two heads have appeared altogether to date?
(d) at least two heads have appeared to date?

3. Let $\mathcal{F}$ and $\mathcal{G}$ be σ-fields of subsets of Ω.
(a) Use elementary set operations to show that $\mathcal{F}$ is closed under countable intersections; that is, if $A_1, A_2, \dots$ are in $\mathcal{F}$, then so is $\bigcap_i A_i$.
(b) Let $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ be the collection of subsets of Ω lying in both $\mathcal{F}$ and $\mathcal{G}$. Show that $\mathcal{H}$ is a σ-field.
(c) Show that $\mathcal{F} \cup \mathcal{G}$, the collection of subsets of Ω lying in either $\mathcal{F}$ or $\mathcal{G}$, is not necessarily a σ-field.

4. Describe the underlying probability spaces for the following experiments:
(a) a biased coin is tossed three times;
(b) two balls are drawn without replacement from an urn which originally contained two ultramarine and two vermilion balls;
(c) a biased coin is tossed repeatedly until a head turns up.

5. Show that the probability that *exactly* one of the events A and B occurs is

$$\mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).$$

6. Prove that $\mathbb{P}(A \cup B \cup C) = 1 - \mathbb{P}(A^{\mathrm{c}} \mid B^{\mathrm{c}} \cap C^{\mathrm{c}})\mathbb{P}(B^{\mathrm{c}} \mid C^{\mathrm{c}})\mathbb{P}(C^{\mathrm{c}})$.

7. (a) If A is independent of itself, show that $\mathbb{P}(A)$ is 0 or 1.
(b) If $\mathbb{P}(A)$ is 0 or 1, show that A is independent of all events B.

8. Let $\mathcal{F}$ be a σ-field of subsets of Ω, and suppose $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies: (i) $\mathbb{P}(\Omega) = 1$, and (ii) $\mathbb{P}$ is additive, in that $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ whenever $A \cap B = \varnothing$. Show that $\mathbb{P}(\varnothing) = 0$.

9. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $B \in \mathcal{F}$ satisfies $\mathbb{P}(B) > 0$. Let $\mathbb{Q} : \mathcal{F} \to [0, 1]$ be defined by $\mathbb{Q}(A) = \mathbb{P}(A \mid B)$. Show that $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space. If $C \in \mathcal{F}$ and $\mathbb{Q}(C) > 0$, show that $\mathbb{Q}(A \mid C) = \mathbb{P}(A \mid B \cap C)$; discuss.

10. Let $B_1, B_2, \dots$ be a partition of the sample space Ω, each B_i having positive probability, and show that

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j).$$

11. Prove **Boole's inequalities**:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i), \qquad \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \ge 1 - \sum_{i=1}^{n} \mathbb{P}(A_i^{\mathrm{c}}).$$

12. Prove that

$$\mathbb{P}\left(\bigcap_{1}^{n} A_i\right) = \sum_{i} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cup A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cup A_j \cup A_k) - \cdots - (-1)^n \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n).$$

13. Let $A_1, A_2, \dots, A_n$ be events, and let N_k be the event that exactly k of the A_i occur. Prove the result sometimes referred to as **Waring's theorem**:

$$\mathbb{P}(N_k) = \sum_{i=0}^{n-k} (-1)^i \binom{k+i}{k} S_{k+i}, \quad \text{where } S_j = \sum_{i_1 < i_2 < \cdots < i_j} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}).$$

Use this result to find an expression for the probability that a purchase of six packets of Corn Flakes yields exactly three distinct busts (see Exercise (1.3.4)).

14. Prove **Bayes's formula**: if $A_1, A_2, \dots, A_n$ is a partition of Ω, each A_i having positive probability, then

$$\mathbb{P}(A_j \mid B) = \frac{\mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}{\sum_1^n \mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}.$$

15. A random number N of dice is thrown. Let A_i be the event that $N = i$, and assume that $\mathbb{P}(A_i) = 2^{-i}$, $i \ge 1$. The sum of the scores is S. Find the probability that:
(a) $N = 2$ given $S = 4$;
(b) $S = 4$ given N is even;
(c) $N = 2$, given that $S = 4$ and the first die showed 1;
(d) the largest number shown by any die is r, where S is unknown.

16. Let $A_1, A_2, \dots$ be a sequence of events. Define

$$B_n = \bigcup_{m=n}^{\infty} A_m, \quad C_n = \bigcap_{m=n}^{\infty} A_m.$$

Clearly $C_n \subseteq A_n \subseteq B_n$. The sequences $\{B_n\}$ and $\{C_n\}$ are decreasing and increasing respectively with limits

$$\lim B_n = B = \bigcap_n B_n = \bigcap_n \bigcup_{m \geq n} A_m, \qquad \lim C_n = C = \bigcup_n C_n = \bigcup_n \bigcap_{m \geq n} A_m.$$

The events B and C are denoted $\limsup_{n\to\infty} A_n$ and $\liminf_{n\to\infty} A_n$ respectively. Show that
(a) $B = \{\omega \in \Omega : \omega \in A_n$ for infinitely many values of $n\}$,
(b) $C = \{\omega \in \Omega : \omega \in A_n$ for all but finitely many values of $n\}$.
We say that the sequence $\{A_n\}$ converges to a limit $A = \lim A_n$ if B and C are the same set A. Suppose that $A_n \to A$ and show that
(c) A is an event, in that $A \in \mathcal{F}$,
(d) $\mathbb{P}(A_n) \to \mathbb{P}(A)$.

17. In Problem (1.8.16) above, show that B and C are independent whenever B_n and C_n are independent for all n. Deduce that if this holds and furthermore $A_n \to A$, then $\mathbb{P}(A)$ equals either zero or one.

18. Show that the assumption that $\mathbb{P}$ is *countably* additive is equivalent to the assumption that $\mathbb{P}$ is continuous. That is to say, show that if a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies $\mathbb{P}(\varnothing) = 0$, $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ whenever $A, B \in \mathcal{F}$ and $A \cap B = \varnothing$, then $\mathbb{P}$ is countably additive (in the sense of satisfying Definition (1.3.1b)) if and only if $\mathbb{P}$ is continuous (in the sense of Lemma (1.3.5)).

19. Anne, Betty, Chloë, and Daisy were all friends at school. Subsequently each of the $\binom{4}{2} = 6$ subpairs meet up; at each of the six meetings the pair involved quarrel with some fixed probability p, or become firm friends with probability $1 - p$. Quarrels take place independently of each other. In future, if any of the four hears a rumour, then she tells it to her firm friends only. If Anne hears a rumour, what is the probability that:
(a) Daisy hears it?
(b) Daisy hears it if Anne and Betty have quarrelled?
(c) Daisy hears it if Betty and Chloë have quarrelled?
(d) Daisy hears it if she has quarrelled with Anne?

20. A biased coin is tossed repeatedly. Each time there is a probability p of a head turning up. Let p_n be the probability that an even number of heads has occurred after n tosses (zero is an even number). Show that $p_0 = 1$ and that $p_n = p(1 - p_{n-1}) + (1 - p)p_{n-1}$ if $n \geq 1$. Solve this difference equation.

21. A biased coin is tossed repeatedly. Find the probability that there is a run of r heads in a row before there is a run of s tails, where r and s are positive integers.

22. A bowl contains twenty cherries, exactly fifteen of which have had their stones removed. A greedy pig eats five whole cherries, picked at random, without remarking on the presence or absence of stones. Subsequently, a cherry is picked randomly from the remaining fifteen.
(a) What is the probability that this cherry contains a stone?
(b) Given that this cherry contains a stone, what is the probability that the pig consumed at least one stone?

23. The '**ménages**' problem poses the following question. Some consider it to be desirable that men and women alternate when seated at a circular table. If n couples are seated randomly according to this rule, show that the probability that nobody sits next to his or her partner is

$$\frac{1}{n!} \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k} \binom{2n-k}{k} (n-k)!$$

You may find it useful to show first that the number of ways of selecting k non-overlapping pairs of adjacent seats is $\binom{2n-k}{k} 2n(2n-k)^{-1}$.

24. An urn contains b blue balls and r red balls. They are removed at random and not replaced. Show that the probability that the first red ball drawn is the $(k+1)$th ball drawn equals $\binom{r+b-k-1}{r-1} \Big/ \binom{r+b}{b}$. Find the probability that the last ball drawn is red.

25. An urn contains a azure balls and c carmine balls, where $ac \neq 0$. Balls are removed at random and discarded until the first time that a ball (B, say) is removed having a different colour from its predecessor. The ball B is now replaced and the procedure restarted. This process continues until the last ball is drawn from the urn. Show that this last ball is equally likely to be azure or carmine.

26. Protocols. A pack of four cards contains one spade, one club, and the two red aces. You deal two cards faces downwards at random in front of a truthful friend. She inspects them and tells you that one of them is the ace of hearts. What is the chance that the other card is the ace of diamonds? Perhaps $\frac{1}{3}$?

Suppose that your friend's protocol was:

(a) with no red ace, say "no red ace",

(b) with the ace of hearts, say "ace of hearts",

(c) with the ace of diamonds but not the ace of hearts, say "ace of diamonds".

Show that the probability in question is $\frac{1}{3}$.

Devise a possible protocol for your friend such that the probability in question is zero.

27. Eddington's controversy. Four witnesses, A, B, C, and D, at a trial each speak the truth with probability $\frac{1}{3}$ independently of each other. In their testimonies, A claimed that B denied that C declared that D lied. What is the (conditional) probability that D told the truth? [This problem seems to have appeared first as a parody in a university magazine of the 'typical' Cambridge Philosophy Tripos question.]

28. The probabilistic method. 10 per cent of the surface of a sphere is coloured blue, the rest is red. Show that, irrespective of the manner in which the colours are distributed, it is possible to inscribe a cube in S with all its vertices red.

29. Repulsion. The event A is said to be repelled by the event B if $\mathbb{P}(A \mid B) < \mathbb{P}(A)$, and to be attracted by B if $\mathbb{P}(A \mid B) > \mathbb{P}(A)$. Show that if B attracts A, then A attracts B, and B^{c} repels A.

If A attracts B, and B attracts C, does A attract C?

30. Birthdays. If m students born on independent days in 1991 are attending a lecture, show that the probability that at least two of them share a birthday is $p = 1 - (365)!/\{(365-m)!\,365^m\}$. Show that $p > \frac{1}{2}$ when $m = 23$.

31. Lottery. You choose r of the first n positive integers, and a lottery chooses a random subset L of the same size. What is the probability that:

(a) L includes no consecutive integers?

(b) L includes exactly one pair of consecutive integers?

(c) the numbers in L are drawn in increasing order?

(d) your choice of numbers is the same as L?

(e) there are exactly k of your numbers matching members of L?

32. Bridge. During a game of bridge, you are dealt at random a hand of thirteen cards. With an obvious notation, show that $\mathbb{P}(4S, 3H, 3D, 3C) \simeq 0.026$ and $\mathbb{P}(4S, 4H, 3D, 2C) \simeq 0.018$. However if suits are not specified, so numbers denote the shape of your hand, show that $\mathbb{P}(4, 3, 3, 3) \simeq 0.11$ and $\mathbb{P}(4, 4, 3, 2) \simeq 0.22$.

33. Poker. During a game of poker, you are dealt a five-card hand at random. With the convention that aces may count high or low, show that:

$$\mathbb{P}(1 \text{ pair}) \simeq 0.423, \quad \mathbb{P}(2 \text{ pairs}) \simeq 0.0475, \quad \mathbb{P}(3 \text{ of a kind}) \simeq 0.021,$$
$$\mathbb{P}(\text{straight}) \simeq 0.0039, \quad \mathbb{P}(\text{flush}) \simeq 0.0020, \quad \mathbb{P}(\text{full house}) \simeq 0.0014,$$
$$\mathbb{P}(4 \text{ of a kind}) \simeq 0.00024, \quad \mathbb{P}(\text{straight flush}) \simeq 0.000015.$$

34. Poker dice. There are five dice each displaying 9, 10, J, Q, K, A. Show that, when rolled:

$$\mathbb{P}(\text{1 pair}) \simeq 0.46, \qquad \mathbb{P}(\text{2 pairs}) \simeq 0.23, \qquad \mathbb{P}(\text{3 of a kind}) \simeq 0.15,$$
$$\mathbb{P}(\text{no 2 alike}) \simeq 0.093, \qquad \mathbb{P}(\text{full house}) \simeq 0.039, \qquad \mathbb{P}(\text{4 of a kind}) \simeq 0.019,$$
$$\mathbb{P}(\text{5 of a kind}) \simeq 0.0008.$$

35. You are lost in the National Park of **Bandrika**†. Tourists comprise two-thirds of the visitors to the park, and give a correct answer to requests for directions with probability $\frac{3}{4}$. (Answers to repeated questions are independent, even if the question and the person are the same.) If you ask a Bandrikan for directions, the answer is always false.

(a) You ask a passer-by whether the exit from the Park is East or West. The answer is East. What is the probability this is correct?

(b) You ask the same person again, and receive the same reply. Show the probability that it is correct is $\frac{1}{2}$.

(c) You ask the same person again, and receive the same reply. What is the probability that it is correct?

(d) You ask for the fourth time, and receive the answer East. Show that the probability it is correct is $\frac{27}{70}$.

(e) Show that, had the fourth answer been West instead, the probability that East is nevertheless correct is $\frac{9}{10}$.

36. Mr Bayes goes to Bandrika. Tom is in the same position as you were in the previous problem, but he has reason to believe that, with probability ϵ, East is the correct answer. Show that:

(a) whatever answer first received, Tom continues to believe that East is correct with probability ϵ,

(b) if the first two replies are the same (that is, either WW or EE), Tom continues to believe that East is correct with probability ϵ,

(c) after three like answers, Tom will calculate as follows, in the obvious notation:

$$\mathbb{P}(\text{East correct} \mid \text{EEE}) = \frac{9\epsilon}{11 - 2\epsilon}, \qquad \mathbb{P}(\text{East correct} \mid \text{WWW}) = \frac{11\epsilon}{9 + 2\epsilon}.$$

Evaluate these when $\epsilon = \frac{9}{20}$.

37. Bonferroni's inequality. Show that

$$\mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) \geq \sum_{r=1}^{n} \mathbb{P}(A_r) - \sum_{r<k} \mathbb{P}(A_r \cap A_k).$$

38. Kounias's inequality. Show that

$$\mathbb{P}\left(\bigcup_{r=1}^{n} A_r\right) \leq \min_k \left\{ \sum_{r=1}^{n} \mathbb{P}(A_r) - \sum_{r:r \neq k} \mathbb{P}(A_r \cap A_k) \right\}.$$

39. The n passengers for a Bell-Air flight in an airplane with n seats have been told their seat numbers. They get on the plane one by one. The first person sits in the wrong seat. Subsequent passengers sit in their assigned seats whenever they find them available, or otherwise in a randomly chosen empty seat. What is the probability that the last passenger finds his seat free?

---

†A fictional country made famous in the Hitchcock film 'The Lady Vanishes'.