Chapter 1

Basic Probability Concepts

1.1 Sample and Event Spaces

1.1.1 Sample Space

A *probabilistic* (or *statistical*) experiment has the following characteristics:

(a) the set of all possible outcomes of the experiment can be described;

(b) the outcome of the experiment cannot be predicted with certainty prior to the performance of the experiment.

The set of all possible outcomes (or sample points) of the experiment is called the *sample space* and is denoted by $\mathcal{S}$. For a given experiment it may be possible to define several sample spaces.

Example For the experiment of tossing a coin three times, we could define

(a) $\mathcal{S} = \{$HHH,HHT,HTH,THH,HTT,THT,TTH,TTT$\}$,
each outcome being an ordered sequence of results; or

(b) $\mathcal{S} = \{0, 1, 2, 3\}$,
each outcome being a possible value for the number of heads obtained.

If $\mathcal{S}$ consists of a list of outcomes (finite or infinite in number), $\mathcal{S}$ is a *discrete sample space*.

Examples

(i) Tossing a die: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$

(ii) Tossing a coin until the first head appears: $\mathcal{S} = \{$H,TH,TTH,TTTH,$\ldots\}$.

Otherwise $\mathcal{S}$ is an *uncountable sample space*. In particular, if $\mathcal{S}$ belongs to a Euclidean space (e.g. real line, plane), $\mathcal{S}$ is a *continuous sample space*.

Example Lifetime of an electronic device: $\mathcal{S} = \{t : 0 \leq t < \infty\}$.

1.1.2 Event Space

Events

A specified collection of outcomes in $\mathcal{S}$ is called an *event*: i.e., any subset of $\mathcal{S}$ (including $\mathcal{S}$ itself) is an event. When the experiment is performed, an event A *occurs* if the outcome is a member of A.

Example In tossing a die once, let the event A be the occurrence of an even number: i.e., $A = \{2, 4, 6\}$. If a 2 or 4 or 6 is obtained when the die is tossed, event A occurs.

The event S is called the *certain event*, since some member of S must occur. A single outcome is called an *elementary event*. If an event contains no outcomes, it is called the *impossible* or *null event* and is denoted by $\emptyset$.

Combination of events

Since events are sets, they may be combined using the notation of set theory: Venn diagrams are useful for exhibiting definitions and results, and you should draw such a diagram for each operation and identity introduced below.

[In the following, $A, B, C, A_1, ..., A_n$ are events in the event space $\mathcal{F}$ (discussed below), and are therefore subsets of the sample space $\mathcal{S}$.]

The *union* of A and B, denoted by $A \cup B$, is the event 'either A or B, or both'.

The *intersection* of A and B, denoted by $A \cap B$, is the event 'both A and B'.

The union and intersection operations are *commutative*, i.e.

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A, \tag{1.1}$$

associative, i.e.

$$A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C, \qquad A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C, \tag{1.2}$$

and *distributive*:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.3}$$

If A is a *subset* of B, denoted by $A \subset B$, then $A \cup B = B$ and $A \cap B = A$.

The *difference* of A and B, denoted by $A \backslash B$, is the event 'A but not B'.

The *complement* of A, denoted by $\overline{A}$, is the event 'not A'.

The complement operation has the properties:

$$A \cup \overline{A} = \mathcal{S}, \quad A \cap \overline{A} = \emptyset, \quad \overline{(\overline{A})} = A, \tag{1.4}$$

and $$\overline{A \cup B} = \overline{A} \cap \overline{B}, \quad \overline{A \cap B} = \overline{A} \cup \overline{B}. \tag{1.5}$$

Note also that $A \backslash B = A \cap \overline{B}$. Hence use of the difference symbol can be avoided if desired (and will be in our discussion).

Two events A and B are termed *mutually exclusive* if $A \cap B = \emptyset$.

Two events A and B are termed *exhaustive* if $A \cup B = \mathcal{S}$.

The above results may be generalised to combinations of n events: thus

$A_1 \cup A_2 \cup ... \cup A_n$ is the event 'at least one of $A_1, A_2, ..., A_n$'

$A_1 \cap A_2 \cap ... \cap A_n$ is the event 'all of $A_1, A_2, ..., A_n$'

$$A_1 \cap (A_2 \cup A_3 \cup ... \cup A_n) = (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup ... \cup (A_1 \cap A_n) \tag{1.6}$$

$$A_1 \cup (A_2 \cap A_3 \cap ... \cap A_n) = (A_1 \cup A_2) \cap (A_1 \cup A_3) \cap ... \cap (A_1 \cup A_n) \tag{1.7}$$

$$\overline{A_1 \cup A_2 \cup ... \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap ... \cap \overline{A_n} \quad \text{or} \quad \overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \overline{A_i} \tag{1.8}$$

$$\overline{A_1 \cap A_2 \cap ... \cap A_n} = \overline{A_1} \cup \overline{A_2} \cup ... \cup \overline{A_n} \quad \text{or} \quad \overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \overline{A_i} \tag{1.9}$$

(The last two results are known as de Morgan's Laws - see e.g. Ross for proofs.)

The events $A_1, A_2, ..., A_n$ are termed *mutually exclusive* if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

The events $A_1, A_2, ..., A_n$ are termed *exhaustive* if $A_1 \cup A_2 \cup ... \cup A_n = \mathcal{S}$.

If the events $A_1, A_2, ..., A_n$ are both mutually exclusive and exhaustive, they are called a *partition* of $\mathcal{S}$.

Event space

The collection of all subsets of S may be too large for probabilities to be assigned reasonably to all its members. This suggests the concept of *event space.*

A collection $\mathcal{F}$ of subsets of the sample space is called an *event space* (or σ-field) if

(a) the certain event $\mathcal{S}$ and the impossible event $\emptyset$ belong to $\mathcal{F}$;

(b) if $A \in \mathcal{F}$, then $\overline{A} \in \mathcal{F}$;

(c) if $A_1, A_2, \ldots \in \mathcal{F}$, then $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$, i.e. $\mathcal{F}$ is closed under the operation of taking countable unions.

(Note that one of the statements in (a) is redundant, in view of (b).)

It is readily shown that, if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. For $\overline{\bigcap_i A_i} = \bigcup_{i=1}^{\infty} (\overline{A_i}) \in \mathcal{F}$ (invoking properties (b) and (c) of $\mathcal{F}$, and the result follows from property (b).

For a *finite* sample space, we normally use the collection of all subsets of $\mathcal{S}$ (the *power set* of $\mathcal{S}$) as the event space. For $\mathcal{S} = \{-\infty, \infty\}$ (or a subset of the real line), the collection of sets containing all one-point sets and all well-defined intervals is an event space.

1.2 Probability Spaces

1.2.1 Probability Axioms

Consideration of the more familiar 'relative frequency' interpretation of probability suggests the following formal definition:

A function P defined on $\mathcal{F}$ is called a *probability function* (or *probability measure*) on $(\mathcal{S}, \mathcal{F})$ if

(a) $0 \leq \mathrm{P}(A) \leq 1$ for all $A \in \mathcal{F}$;

(b) $\mathrm{P}(\emptyset) = 0$ and $\mathrm{P}(\mathcal{S}) = 1$;

(c) if $A_1, A_2, \ldots$ is a collection of mutually exclusive events $\in \mathcal{F}$, then

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots \qquad (1.10)$$

Note:

(i) There is redundancy in both axioms (a) and (b) - see Corollaries to the Complementarity Rule (overleaf).

(ii) Axiom (c) states that P is *countably additive.* This is a more extensive property than being *finitely additive*, i.e. the property that, if $A_1, A_2, \ldots A_n$ is a collection of n mutually exclusive events, then

$$\mathrm{P}(A_1 \cup A_2 \cup \ldots A_n) = \sum_{i=1}^{n} \mathrm{P}(A_i). \qquad (1.11)$$

This is derived from axiom (c) by defining $A_{n+1}, A_{n+2}, \ldots$ to be $\emptyset$: only (1.11) is required in the case of *finite* $\mathcal{S}$ and $\mathcal{F}$. It can be shown that the property of being countably additive is equivalent to the assertion that P is a *continuous set function*, i.e. given an increasing sequence of events $A_1 \subseteq A_2 \subseteq \ldots$ and writing $A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i$, then $\mathrm{P}(A) = \lim_{i \to \infty} \mathrm{P}(A_i)$ (with a similar result for a decreasing sequence) - see e.g. Grimmett & Welsh §1.9

The triple $(\mathcal{S}, \mathcal{F}, \mathrm{P})$ is called a *probability space.*

The function P(.) assigns a numerical value to each $A \in \mathcal{F}$; however, the above axioms (a)-(c) do not tell us *how* to assign the numerical value – this will depend on the description of, and assumptions concerning, the experiment in question.

If $\mathrm{P}(A) = a/(a+b)$, the *odds* for A occurring are a to b and the odds against A occurring are b to a.

1.2.2 Derived Properties

From the probability axioms (a)-(c), many results follow. For example:

(i) Complementarity rule If $A \in \mathcal{F}$, then

$$P(A) + P(\overline{A}) = 1. \tag{1.12}$$

Proof $\mathcal{S} = A \cup \overline{A}$ - the union of two disjoint (or mutually exclusive) events.
So $P(\mathcal{S}) = 1 = P(A) + P(\overline{A})$ by probability axiom (c). ♢

Corollaries

(a) Let $A = \mathcal{S}$. Then $P(\mathcal{S}) + P(\emptyset) = 1$, so

$$P(\mathcal{S}) = 1 \Rightarrow P(\emptyset) = 0$$

so axiom (b) can be simplified to $P(\mathcal{S}) = 1$.

(b) If $A \subset \mathcal{S}$, then $P(\overline{A}) \geq 0 \Rightarrow P(A) \leq 1$, so in axiom (a) only $P(A) \geq 0$ is needed.

(ii)

$$\text{If } \quad A \subset B \text{ , then } \quad P(A) \leq P(B). \tag{1.13}$$

Proof $P(B) = P(A \cup (B \cap \overline{A})) = P(A) + P(B \cap \overline{A})$ using axiom (c). The result follows from the fact that $P(B \cap \overline{A}) \geq 0$ (axiom (a)). ♢

(iii) Addition Law (for two events) If $A, B \in \mathcal{F}$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.14}$$

Proof We have that

$$\begin{aligned} A \cup B &= A \cup (B \cap \overline{A}) \\ \text{and} \qquad B &= (A \cap B) \cup (B \cap \overline{A}). \end{aligned}$$

In each case, the r.h.s. is the union of two mutually exclusive events $\in F$. So by axiom (c)

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \cap \overline{A}) \\ \text{and} \qquad P(B) &= P(A \cap B) + P(B \cap \overline{A}). \end{aligned}$$

Eliminating $P(B \cap \overline{A})$, the result follows. ♢

Corollaries Boole's Inequalities If $A, B \in \mathcal{F}$, then it follows from (1.14) and axiom (a) that

(a) $P(A \cup B) \leq P(A) + P(B);$ (1.15a)

(b) $P(A \cap B) \geq P(A) + P(B) - 1.$ (1.15b)

(iv) Generalized Addition Law (for n events) If $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then the probability that at least one of these events occurs is

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \ldots\ldots\ldots \\ &+ (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots A_n). \end{aligned} \tag{1.16}$$

This is sometimes termed the *inclusion-exclusion principle*. One *proof* is by induction on n: this is left as an exercise. For a non-inductive proof, see Ross.

(v) Bonferroni's Inequalities

If $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then

$$\text{(a)} \quad \sum_{i=1}^{n} \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i \cap A_j) \;\le\; \mathrm{P}(A_1 \cup \ldots \cup A_n) \le \sum_{i=1}^{n} \mathrm{P}(A_i); \tag{1.17a}$$

$$\text{(b)} \quad \mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_n) \;\ge \sum_{i=1}^{n} \mathrm{P}(A_i) - (n-1) \;=\; 1 - \sum_{i=1}^{n} \mathrm{P}(\overline{A}_i). \tag{1.17b}$$

Again, *proofs* are by induction on n (see Examples Sheet 1). [(1.17a) can be generalized.]

1.3 Solving Probability Problems

The axioms and derived properties of P provide a probability 'calculus', but they do not help in setting up a probability model for an experiment or in assigning actual numerical values to the probabilities of specific events. However, the properties of P may prove very useful in solving a problem once these two steps have been completed.

An important special case (the 'classical' situation) is where $\mathcal{S}$ is finite and consists of N equally likely outcomes $E_1, E_2, \ldots, E_N$: thus, $E_1, \ldots, E_N$ are mutually exclusive and exhaustive, and $\mathrm{P}(E_1) = \ldots \mathrm{P}(E_N)$. Then, since

$$1 = \mathrm{P}(\mathcal{S}) = \mathrm{P}(E_1 \cup \ldots \cup E_N) = \mathrm{P}(E_1) + \cdots + \mathrm{P}(E_N),$$

we have

$$\mathrm{P}(E_i) = \frac{1}{N}, \qquad i = 1, \ldots, N. \tag{1.18}$$

Often there is more than one way of tackling a probability problem: on the other hand, an approach which solves one problem may be inapplicable to another problem (or only solve it with considerable difficulty). There are five general methods which are widely used:

(i) **Listing** the elements of the sample space $\mathcal{S} = \{E_1, \ldots, E_N\}$, where N is finite (and fairly small), and identifying the 'favourable' ones. Thus, if the event of interest is $A = \{E_{i_1}, \ldots, E_{i_{N(A)}}\}$, then

$$\mathrm{P}(A) = \mathrm{P}(E_{i_1}) + \cdots + \mathrm{P}(E_{i_{N(A)}}).$$

If the outcomes can be assumed to be equally likely, this reduces to

$$\mathrm{P}(A) = \frac{N(A)}{N}; \tag{1.19}$$

i.e.

$$\mathrm{P}(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}.$$

In such calculations, *symmetry* arguments may play an important role.

(ii) **Enumeration**, when S is a finite sample space of equally likely outcomes. To avoid listing, we calculate the numbers $N(A)$ and N by combinatorial arguments, then again use $\mathrm{P}(A) = N(A)/N$.

(iii) **Sequential** (or **Historical**) method. Here we follow the history of the system (actual or conceptual) and use the multiplication rules for probabilities (see later).

(iv) **Recursion**. Here we express the problem in some recursive form, and solve it by recursion or induction (see later).

(v) Define suitable **random variables** e.g. **indicator** random variables (see later).

In serious applications, we should also investigate the *robustness* (or sensitivity of the results) with respect to small changes in the initially assumed probabilities and assumptions in the model (e.g. assumed independence of successive trials).

1.4 Some Results in Combinatorial Analysis

As already observed, in many problems involving finite sample spaces with equally likely outcomes, the calculation of a probability reduces to problems of *combinatorial counting*. Also, it is common for problems which at first appear to be quite different to actually have identical solutions: e.g. (see §1.6.2 below)

(i) calculate the probability that all students in a class of 30 have different birthdays;

(ii) calculate the probability that in a random sample of size 30, sampling with replacement, from a set of 365 distinct objects, all objects are different.

In this course, we will not be doing many such examples, but from time to time combinatorial arguments will be used, and you will be expected to be familiar with the more elementary results in combinatorial analysis. A summary of the different contexts in which these can arise is provided below.

All the following results follow from simple counting principles, and it is worth checking that you can derive them.

Sampling

Suppose that a sample of size r is taken from a population of n distinct elements. Then

(i) If sampling is *with replacement*, there are n^r different *ordered* samples.

(ii) If sampling is *without replacement*, there are $n!/(n-r)!$ different *ordered* samples and $\binom{n}{r} = n!/\{r!(n-r)!\}$ different *non-ordered* samples.

(A *random sample* of size r from a finite population is one which has been drawn in such a way that all possible samples of size r have the same probability of being chosen. If a sample is drawn by successively selecting elements from the population (with or without replacement) in such a way that each element remaining has the same chance of being chosen at the next selection, then the sample is a random sample.)

Arrangements

(i) The number of different arrangements of r objects chosen from n distinct objects is $n!/(n-r)!$. In particular, the number of arrangements of n distinct objects is $n!$.

(ii) The number of different arrangements of n objects of which n_1 are of one kind, n_2 of a second kind,...,n_k of a kth kind is $n!/\{n_1!n_2!...n_k!\}$, where $\sum_{i=1}^{k} n_i = n$ (known as the *multinomial coefficient*). In particular, when $k = 2$, this reduces to $\binom{n}{n_1}$, the *binomial coefficient*.

Sub-populations

(i) The number of ways in which a population of n distinct elements can be partitioned into k sub-populations with n_1 elements in the first, n_2 elements in the second,..., n_k elements in the kth, so that $\sum_{i=1}^{k} n_i = n$, is the multinomial coefficient $n!/\{n_1!n_2!...n_k!\}$.

(ii) Consider a population of n distinct elements which is partitioned into k sub-populations, where the ith sub-population contains n_i elements. Then the number of different *non-ordered* samples of size r, sampling *without replacement*, such that the sample contains r_1 elements from the first sub-population, r_2 elements from the second,..., r_k elements from the kth, is

$$\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_k}{r_k}, \qquad \sum_{i=1}^{k} n_i = n, \quad \sum_{i=1}^{k} r_i = r$$

1.5 Matching and Collection Problems

1.5.1 Matching Problems

Example 1.1

Suppose n cards numbered $1, 2, ...n$ are laid out at random in a row. Define the event

A_i: card i appears in the ith position of the row.

This is termed a *match* (or *coincidence* or *rencontre*) in the ith position. What is the probability of obtaining at least one match?

Solution First, some preliminary results:

(a) How many arrangements are there

in all? — $n!$.

with a match in position 1? — $(n-1)!$ (by ignoring position 1).

with a match in position i? — $(n-1)!$ (by symmetry).

with matches in positions 1 and 2? — $(n-2)!$

with matches in positions i and j? — $(n-2)!$

with matches in positions i, j, k? — $(n-3)!$

etc.

(b) How many terms are there in the summations

$$\sum_{1\le i<j\le n} b_{ij}, \qquad \sum_{1\le i<j<k\le n} c_{ijk}, \qquad ....?$$

The easiest argument is as follows. In the case of the first summation, take a sample of size 2, {r,s} say, from the population $\{1, 2, \dots, n\}$, sampling without replacement: let $i = \min(r, s), j = \max(r, s)$. The number of such samples is $\binom{n}{2}$, so this is the number of terms in the first summation.

Similarly, the number of terms in the second summation is $\binom{n}{3}$ – and so on.

Now for the main problem. We have that

$$\begin{aligned}\text{P(at least one match)} &= \text{P}(A_i \cup A_2 \cup \ldots \cup A_n) \\ &= \sum_{i=1}^{n} \text{P}(A_i) - \sum_{i<j} \text{P}(A_i \cap A_j) + \cdots + (-1)^{n+1}\text{P}(A_1 \cap \cdots \cap A_n)\end{aligned}$$

by the generalized addition law (1.16). From (a) above,

$$\begin{aligned}\text{P}(A_i) &= \frac{(n-1)!}{n!}, \qquad i = 1, \ldots, n \\ \text{P}(A_i \cup A_j) &= \frac{(n-2)!}{n!}, \qquad 1 \leq i < j \leq n\end{aligned}$$

and more generally

$$\text{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) = \frac{(n-r)!}{n!}, \qquad 1 \leq i_1 < i_2 < \cdots < i_r \leq n.$$

From preliminary result (b) above, the number of terms in the summations are

$$n, \binom{n}{2}, \binom{n}{3}, \ldots, \binom{n}{r}, \ldots, \binom{n}{n} = 1$$

respectively. Hence

$$\begin{aligned}\text{P}(A_1 \cup \cdots \cup A_n) &= n.\frac{(n-1)!}{n!} - \binom{n}{2}.\frac{(n-2)!}{n!} + \binom{n}{3}.\frac{(n-3)!}{n!} - \cdots + (-1)^{n+1}\binom{n}{n}\frac{0!}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1}.\frac{1}{n!}.\end{aligned} \tag{1.20}$$

Now

$$e^x = \sum_{r=0}^{\infty} \frac{x^r}{r!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \qquad \text{for all real } x.$$

So

$$e^{-1} = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots,$$

and therefore

$$\text{P}(A_1 \cup \ldots \cup A_n) \approx 1 - e^{-1} = 0.632121 \text{ when } n \text{ is large.} \tag{1.21}$$

(The approximation may be regarded as adequate for $n \geq 6$: for $n = 6$, the exact result is 0.631944.) We deduce also that

$$\text{P}(\textit{no} \text{ match}) \approx e^{-1} = 0.367879. \tag{1.22}$$

An outcome in which there is no match is termed a *derangement*.

We shall return to this problem from time to time. The problem can be extended to finding the probability of exactly k matches $(0 \leq k \leq n)$ – see Examples Sheet 1.

This matching problem is typical of many probability problems in that it can appear in a variety of guises: e.g.

matching of two packs of playing cards (consider one pack to be in a fixed order);

envelopes and letters are mixed up by accident and then paired randomly;

n drivers put their car keys into a basket, then each driver draws a key at random.

It is important to be able to recognize the essential equivalence of such problems.

1.5.2 Collecting Problems

Example 1.2 *A Token-Collecting Problem* (or *Adding Spice to Breakfast*)

In a promotional campaign, a company making a breakfast cereal places one of a set of 4 cards in each packet of cereal. Anyone who collects a complete set of cards can claim a prize. Assuming that the cards are distributed at random into the packets, calculate the probability that a family which purchases 10 packets will be able to claim at least one prize.

Solution Define

A_i: family does *not* find card i ($i = 1, \ldots, 4$) in its purchase of 10 packets.

Then the required probability is

$$\mathrm{P}(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3 \cap \overline{A}_4) = 1 - \mathrm{P}(\overline{\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3 \cap \overline{A}_4}) = 1 - \mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4).$$

Again invoking the generalized addition law (1.16), we have

$$\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4) = \sum_{i=1}^{4} \mathrm{P}(A_i) - \sum_{i<j} \mathrm{P}(A_i \cap A_j) + \sum_{i<j<k} \mathrm{P}(A_i \cap A_j \cap A_k) - \mathrm{P}(A_1 \cap A_2 \cap A_3 \cap A_4).$$

Now

$$\mathrm{P}(A_1) = \mathrm{P}(\text{no card 1 in 10 packets}) = \left(\frac{3}{4}\right)^{10}$$

(by combinatorial argument $\dfrac{3 \times 3 \times \ldots \times 3}{4 \times 4 \times \ldots \times 4}$ or multiplication law or binomial distribution (see later)). By symmetry

$$\mathrm{P}(A_i) = \left(\frac{3}{4}\right)^{10}, \qquad i = 1, \ldots, 4.$$

Similarly, we have that

$$\begin{aligned} \mathrm{P}(A_i \cap A_j) &= \left(\frac{2}{4}\right)^{10}, \qquad i \neq j, \\ \mathrm{P}(A_i \cap A_j \cap A_k) &= \left(\frac{1}{4}\right)^{10}, \qquad i \neq j \neq k, \\ \mathrm{P}(A_1 \cap A_2 \cap A_3 \cap A_4) &= 0. \end{aligned}$$

So

$$\mathrm{P}(A_1 \cup A_2 \cup A_3 \cup A_4) = 4\left(\frac{3}{4}\right)^{10} - 6\left(\frac{2}{4}\right)^{10} + 4\left(\frac{1}{4}\right)^{10} = 0.2194.$$

Hence the required probability is $1 - 0.2194 \approx 0.78$. ♢

1.6 Further Examples

Example 1.3 *A Sampling Problem*

A bag contains 12 balls numbered 1 to 12. A random sample of 5 balls is drawn (without replacement). What is the probability that

(a) the largest number chosen is 8;

(b) the median number in the sample is 8?

Solution There are $\binom{12}{5}$ points in the sample space. Then:

(a) If the largest number chosen is 8, the other 4 are chosen from the numbers 1 to 7 and the number of ways this can be done is $\binom{7}{4}$; so the required probability is

$$\frac{\binom{7}{4}}{\binom{12}{5}}.$$

(b) If the median is 8, two of the chosen numbers come from $\{1, 2, 3, 4, 5, 6, 7\}$ and 2 from $\{9, 10, 11, 12\}$; so the required probability is

$$\frac{\binom{7}{2} \cdot \binom{4}{2}}{\binom{12}{5}}. \qquad \square$$

Example 1.4 *The Birthday Problem*

If there are n people in a room, what is the probability that at least two of them have the same birthday?

Solution Recording the n birthdays is equivalent to taking a sample of size n with replacement from a population of size 365 (leap-years will be ignored). The number of samples is $(365)^n$ and we assume that each is equally likely (corresponds to the assumption that birth rates are constant throughout the year). The number of samples in which there are *no* common birthdays is clearly

$$365 \times 364 \times ... \times (365 - (n-1)).$$

So the probability that there are *no* common birthdays is

$$\frac{(365)(364)...(365-(n-1))}{365^n} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) ... \left(1 - \frac{n-1}{365}\right). \qquad (1.23)$$

The required probability is 1 - this expression: it increases rapidly with n, first exceeding $\frac{1}{2}$ at $n = 23$ (value 0.5073). $\diamondsuit$

[Some people may find this result surprising, thinking that the critical value of n should be much larger; but this is perhaps because they are confusing this problem with another one - what is the probability that at least one person in a group of n people has a *specific* birthday e.g. 1st January. The answer to this is clearly $1-(\frac{364}{365})^n$: this rises more slowly with n and does not exceed $\frac{1}{2}$ until $n = 253$.]

Example 1.5 *A Seating Problem*

If n married couples are seated at random at a round table, find the probability that no wife sits next to her husband.

Solution Let A_i: couple i sit next to each other. Then the required probability is

$$\begin{aligned} & 1-\mathrm{P}(A_1 \cup A_2 \cup \ldots \cup A_n) \\ = \; & 1-\sum_{i=1}^{n} \mathrm{P}(A_i) - \cdots - (-1)^{r+1} \sum_{1 \le i_1 < i_2 < \ldots < i_r \le n} \mathrm{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) - \cdots \\ & -(-1)^{n+1}\mathrm{P}(A_1 \cap \cdots \cap A_n). \end{aligned}$$

To compute the generic term

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r})$$

we proceed as follows. There are $(2n-1)!$ ways of seating $2n$ people at a round table (put the first person on some seat, then arrange the other $(2n-1)$ around them). The number of arrangements in which a specified set of r men, $i_1, i_2, ..., i_r$, sit next to their wives is $(2n-r-1)!$ (regard each specified couple as a single entity - then we have to arrange $(2n-2r+r) = (2n-r)$ entities (the r specified couples and the other people) around the table). Also, each of the r specified couples can be seated together in two ways. So the number of arrangements in which a specified set of r men sit next to their wives is $2^r(2n-r-1)!$. Hence

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) = \frac{2^r(2n-r-1)!}{(2n-1)!}.$$

It follows that the required probability is

$$1-\sum_{r=1}^{n}(-1)^{r+1}\binom{n}{r}2^r(2n-r-1)!/(2n-1)!. \qquad \square$$

(Note: there is another version of this problem in which one imposes the constraint that men and women must alternate around the table).