1 DEFINITIONS, TERMINOLOGY, NOTATION

1.1 EVENTS AND THE SAMPLE SPACE

- An **experiment** is a one-off or repeatable process or procedure for which
  (a) there is a well-defined set of (possible) outcomes
  (b) the actual outcome is not known with certainty.
- A **sample outcome**, \( \omega \), is precisely one of the (possible) outcomes of an experiment.
- The **sample space**, \( \Omega \), of an experiment is the set of all (possible) outcomes.

\( \Omega \) is a set in the mathematical sense, so set theory notation can be used. For example, if the sample outcomes are denoted \( \omega_1, \omega_2, \ldots \), say, then the sample space of an experiment can be

- a **finite** list of sample outcomes, \( \{\omega_1, \ldots, \omega_k\} \)
- an **infinite** list of sample outcomes, \( \{\omega_1, \omega_2, \ldots\} \)
- an **interval** or **region** of a real space, \( \{\omega : \omega \in \mathbb{A} \subseteq \mathbb{R}^d\} \)

An **event**, \( E \), is a designated collection of sample outcomes. Event \( E \) **occurs** if the actual outcome of the experiment is one of this collection; for any event \( E, E \subseteq \Omega \).

- the collection of **all** sample outcomes, \( \Omega \),
- the collection of **none** of the sample outcomes, \( \emptyset \) (the **empty set**).

1.1.1 OPERATIONS IN SET THEORY

Set theory operations can be used to manipulate events in probability theory. Consider events \( E, F \subseteq \Omega \). Then the three basic operations are

<table>
<thead>
<tr>
<th>Operation</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union</td>
<td>( E \cup F )</td>
<td>( E ) or ( F ) both occur</td>
</tr>
<tr>
<td>Intersection</td>
<td>( E \cap F )</td>
<td>both ( E ) and ( F ) occur</td>
</tr>
<tr>
<td>Complement</td>
<td>( E' )</td>
<td>( E ) does not occur</td>
</tr>
</tbody>
</table>

Consider events \( E, F, G \subseteq \Omega \).

<table>
<thead>
<tr>
<th>Property</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commutativity</td>
<td>( E \cup F = F \cup E )</td>
</tr>
<tr>
<td></td>
<td>( E \cap F = F \cap E )</td>
</tr>
<tr>
<td>Associativity</td>
<td>( E \cup (F \cup G) = (E \cup F) \cup G )</td>
</tr>
<tr>
<td></td>
<td>( E \cap (F \cap G) = (E \cap F) \cap G )</td>
</tr>
<tr>
<td>Distributivity</td>
<td>( E \cup (F \cap G) = (E \cup F) \cap (E \cup G) )</td>
</tr>
<tr>
<td></td>
<td>( E \cap (F \cup G) = (E \cap F) \cup (E \cap G) )</td>
</tr>
<tr>
<td>De Morgan's Laws</td>
<td>( (E \cup F)' = E' \cap F' )</td>
</tr>
<tr>
<td></td>
<td>( (E \cap F)' = E' \cup F' )</td>
</tr>
</tbody>
</table>

Union and intersection are **binary** operators, that is, they take only two arguments, and thus the bracketing in the above equations is necessary. For \( k \geq 2 \) events, \( E_1, E_2, \ldots, E_k \),

\[
\bigcup_{i=1}^{k} E_i = E_1 \cup \ldots \cup E_k \quad \text{and} \quad \bigcap_{i=1}^{k} E_i = E_1 \cap \ldots \cap E_k
\]

for the union and intersection of \( E_1, E_2, \ldots, E_k \), with a further extension for \( k \) **infinite**.
1.1.2 MUTUALLY EXCLUSIVE EVENTS AND PARTITIONS

Events $E$ and $F$ are **mutually exclusive** if $E \cap F = \emptyset$, that is, if events $E$ and $F$ cannot both occur. If the sets of sample outcomes represented by $E$ and $F$ are **disjoint** (have no common element), then $E$ and $F$ are mutually exclusive.

Events $E_1, \ldots, E_k \subseteq \Omega$ form a **partition** of event $F \subseteq \Omega$ if

(a) $E_i \cap E_j = \emptyset$ for $i \neq j, i, j = 1, \ldots, k$

(b) $\bigcup_{i=1}^k E_i = F$

so that each element of the collection of sample outcomes corresponding to event $F$ is in **one and only one** of the collections corresponding to events $E_1, \ldots, E_k$.

1.1.3 SIGMA-ALGEBRAS

A (countable) collection of subsets, $\mathcal{E}$, of sample space $\Omega$, say $\mathcal{E} = \{E_1, E_2, \ldots\}$, is a **sigma-algebra** if

I $\Omega \in \mathcal{E}$

II $E \in \mathcal{E} \implies E' \in \mathcal{E}$

III If $E_1, E_2, \ldots \in \mathcal{E}$, then $\bigcup_{i=1}^\infty E_i \in \mathcal{E}$.

If $\mathcal{E}$ is an algebra of subsets of $\Omega$, then

(i) $\emptyset \in \mathcal{E}$

(ii) If $E_1, E_2 \in \mathcal{E}$, then $E_1', E_2' \in \mathcal{E}$.

so $\mathcal{E}$ is also **closed under intersection**.

1.2 THE PROBABILITY FUNCTION

For an event $E \subseteq \Omega$, the **probability that $E$ occurs** is written $P(E)$.

**Interpretation** : $P(.)$ is a **set-function** that assigns “weight” to collections of possible outcomes of an experiment. There are many ways to think about precisely how this assignment is achieved;

**CLASSICAL** : “Consider equally likely sample outcomes ...”

**FREQUENTIST** : “Consider long-run relative frequencies ...”

**SUBJECTIVE** : “Consider personal degree of belief ...”

or merely think of $P(.)$ as a set-function.
1.3 PROPERTIES OF $P(.)$: THE AXIOMS OF PROBABILITY

Consider sample space $\Omega$. Then probability function $P(.)$ acts on a sigma-algebra $\mathcal{E}$ defined on $\Omega$

$$P: \mathcal{E} \rightarrow \mathbb{R}$$

and satisfies the following properties:

(I) Let $E \in \mathcal{E}$. Then $0 \leq P(E) \leq 1$.

(II) $P(\Omega) = 1$.

(III) If $E_1, E_2, \ldots$ are mutually exclusive events, then

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

1.3.1 COROLLARIES TO THE PROBABILITY AXIOMS

For events $E, F \subseteq \Omega$

1. $P(E') = 1 - P(E)$, and hence $P(\emptyset) = 0$.

2. If $E \subseteq F$, then $P(E) \leq P(F)$.

3. In general, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

4. $P(E \cap F') = P(E) - P(E \cap F)$.

5. $P(E \cup F) \leq P(E) + P(F)$.

6. $P(E \cap F) \geq P(E) + P(F) - 1$.

The general addition rule for probabilities and Boole’s Inequality extend to more than two events. Let $E_1, \ldots, E_n$ be events in $\Omega$. Then

(i) $P \left( \bigcup_{i=1}^{n} E_i \right) \leq \sum_{i=1}^{n} P(E_i)$.

(ii) $P \left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i} P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k) - \ldots + (-1)^{n+1} P \left( \bigcap_{i=1}^{n} E_i \right)$

(i) follows from 5; for (ii), construct the events $F_1 = E_1$ and

$$F_i = E_i \cap \left( \bigcup_{k=1}^{i-1} E_k \right)'$$

for $i = 2, 3, \ldots, n$. Then $F_1, F_2, \ldots, F_n$ are disjoint, and $\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} F_i$, so

$$P \left( \bigcup_{i=1}^{n} E_i \right) = P \left( \bigcup_{i=1}^{n} F_i \right) = \sum_{i=1}^{n} P(F_i).$$

Now, by the corollary above, for $i = 2, 3, \ldots, n$,

$$P(F_i) = P(E_i) - P \left( E_i \cap \left( \bigcup_{k=1}^{i-1} E_k \right) \right) = P(E_i) - P \left( \bigcup_{k=1}^{i-1} (E_i \cap E_k) \right)$$

and the result follows by recursive expansion of the second term for $i = 2, 3, \ldots, n$. 

3
1.4 CONDITIONAL PROBABILITY

For events $E, F \subseteq \Omega$ the conditional probability that $F$ occurs given that $E$ occurs is written $P(F|E)$, and is defined by

$$P(F|E) = \frac{P(E \cap F)}{P(E)}$$

if $P(E) > 0$.

**NOTE:** $P(E \cap F) = P(E)P(F|E)$, and in general, for events $E_1, \ldots, E_k$,

$$P \left( \bigcap_{i=1}^{k} E_i \right) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \ldots P(E_k|E_1 \cap E_2 \cap \ldots \cap E_{k-1}).$$

This result is known as the chain or multiplication rule.

Events $E$ and $F$ are independent if

$$P(E|F) = P(E)$$

so that $P(E \cap F) = P(E)P(F)$.

**Extension:** Events $E_1, \ldots, E_k$ are independent if, for every subset of events of size $l \leq k$, indexed by \{i_1, \ldots, i_l\}, say,

$$P \left( \bigcap_{j=1}^{l} E_{i_j} \right) = \prod_{j=1}^{l} P(E_{i_j}).$$

1.5 THE THEOREM OF TOTAL PROBABILITY

**THEOREM**

Let $E_1, \ldots, E_k$ be a partition of $\Omega$, and let $F \subseteq \Omega$. Then

$$P(F) = \sum_{i=1}^{k} P(F|E_i)P(E_i)$$

**PROOF**

$E_1, \ldots, E_k$ form a partition of $\Omega$, and $F \subseteq \Omega$, so

$$F = (F \cap E_1) \cup \ldots \cup (F \cap E_k)$$

$$\Rightarrow P(F) = \sum_{i=1}^{k} P(F \cap E_i) = \sum_{i=1}^{k} P(F|E_i)P(E_i)$$

3, as $E_i \cap E_j = \emptyset$).

**Extension:** If we assume that Axiom 3 holds, that is, that $P$ is countably additive, then the theorem still holds, that is, if $E_1, E_2, \ldots$ are a partition of $\Omega$, and $F \subseteq \Omega$, then

$$P(F) = \sum_{i=1}^{\infty} P(F \cap E_i) = \sum_{i=1}^{\infty} P(F|E_i)P(E_i)$$

if $P(E_i) > 0$ for all $i$. 

4
1.6 BAYES THEOREM

THEOREM

Suppose \( E, F \subseteq \Omega \), with \( P(E)P(F) > 0 \). Then

\[
P(E|F) = \frac{P(F|E)P(E)}{P(F)}
\]

PROOF

\[
P(E|F)P(F) = P(E \cap F) = P(F|E)P(E), \quad \text{so} \quad P(E|F)P(F) = P(F|E)P(E).
\]

Extension: If \( E_1, \ldots, E_k \) are disjoint, with \( P(E_i) > 0 \) for \( i = 1, \ldots, k \), and form a partition of \( F \subseteq \Omega \), then

\[
P(E_i|F) = \frac{P(F|E_i)P(E_i)}{\sum_{i=1}^{k} P(F|E_i)P(E_i)}
\]

The extension to the countably additive (infinite) case also holds.

NOTE: in general, \( P(E|F) \neq P(F|E) \)

1.7 COUNTING TECHNIQUES

Suppose that an experiment has \( N \) equally likely sample outcomes. If event \( E \) corresponds to a collection of sample outcomes of size \( n(E) \), then

\[
P(E) = \frac{n(E)}{N}
\]

so it is necessary to be able to evaluate \( n(E) \) and \( N \) in practice.

1.7.1 THE MULTIPLICATION PRINCIPLE

If operations labelled \( 1, \ldots, r \) can be carried out in \( n_1, \ldots, n_r \) ways respectively, then there are

\[
\prod_{i=1}^{r} n_i = n_1 \times \ldots \times n_r
\]

ways of carrying out the \( r \) operations in total.

Example 1.1 If each of \( r \) trials of an experiment has \( N \) possible outcomes, then there are \( N^r \) possible sequences of outcomes in total. For example:

(i) If a multiple choice exam has 20 questions, each of which has 5 possible answers, then there are \( 5^{20} \) different ways of completing the exam.

(ii) There are \( 2^m \) subsets of \( m \) elements (as each element is either in the subset, or not in the subset, which is equivalent to \( m \) trials each with two outcomes).


1.7.2 SAMPLING FROM A FINITE POPULATION

Consider a collection of \( N \) items, and a sequence of operations labelled \( 1, \ldots, r \) such that the \( i \)th operation involves selecting one of the items remaining after the first \( i - 1 \) operations have been carried out. Let \( n_i \) denote the number of ways of carrying out the \( i \)th operation, for \( i = 1, \ldots, r \). Then

(a) **Sampling with replacement**: an item is returned to the collection after selection. Then \( n_i = N \) for all \( i \), and there are \( N^r \) ways of carrying out the \( r \) operations.

(b) **Sampling without replacement**: an item is not returned to the collection after selected. Then \( n_i = N - i + 1 \), and there are \( N(N - 1)\ldots(N - r + 1) \) ways of carrying out the \( r \) operations.

e.g. Consider selecting 5 cards from 52. Then

(a) leads to \( 52^5 \) possible selections, whereas

(b) leads to \( 52 \times 51 \times 50 \times 49 \times 48 \) possible selections

**NOTE**: The order in which the operations are carried out may be important e.g. in a raffle with three prizes and 100 tickets, the draw \{45, 19, 76\} is different from \{19, 76, 45\}.

**NOTE**: The items may be **distinct** (unique in the collection), or **indistinct** (of a unique type in the collection, but not unique individually). For example, the numbered balls in a lottery, or individual playing cards, are **distinct**. However balls in the lottery are regarded as “WINNING” or “NOT WINNING”, or playing cards are regarded in terms of their suit only, are **indistinct**.

1.7.3 PERMUTATIONS AND COMBINATIONS

- A **permutation** is an ordered arrangement of a set of items.
- A **combination** is an unordered arrangement of a set of items.

**RESULT 1** The number of permutations of \( n \) distinct items is \( n! = n(n - 1)\ldots 1 \).

**RESULT 2** The number of permutations of \( r \) from \( n \) distinct items is

\[
P_r^n = \frac{n!}{(n-r)!} = n(n-1)\ldots(n-r+1) \quad \text{(by the Multiplication Principle)}.
\]

If the order in which items are selected is not important, then

**RESULT 3** The number of combinations of \( r \) from \( n \) distinct items is

\[
C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!} \quad \text{(as } P_r^n = r!C_r^n \text{)}.
\]

- recall the **Binomial Theorem**, namely

\[
(a + b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.
\]

Then the number of subsets of \( m \) items can be calculated as follows; for each \( 0 \leq j \leq m \), choose a subset of \( j \) items from \( m \). Then

\[
\text{Total number of subsets} = \sum_{j=0}^{m} \binom{m}{j} = (1 + 1)^m = 2^m.
\]

If the items are **indistinct**, but each is of a unique type, say Type I, \ldots, Type \( \kappa \) say, (the so-called **Urn Model**) then, then a more general formula applies:
RESULT 4 The number of distinguishable permutations of \( n \) indistinct objects, comprising \( n_i \) items of type \( i \) for \( i = 1, \ldots, \kappa \) is

\[
\frac{n!}{n_1! n_2! \ldots n_\kappa!}
\]

Special Case: if \( \kappa = 2 \), then the number of distinguishable permutations of the \( n_1 \) objects of type I, and \( n_2 = n - n_1 \) objects of type II is

\[
C_{n_2}^n = \frac{n!}{n_1!(n-n_1)!}
\]

Also, there are \( C_r^n \) ways of partitioning \( n \) distinct items into two “cells”, with \( r \) in one cell and \( n - r \) in the other.

1.7.4 PROBABILITY CALCULATIONS

Recall that if an experiment has \( N \) equally likely sample outcomes, and event \( E \) corresponds to a collection of sample outcomes of size \( n(E) \), then

\[
P(E) = \frac{n(E)}{N}
\]

**Example 1.2** A True/False exam has 20 questions. Let \( E = \) “16 answers correct at random”. Then

\[
P(E) = \frac{\text{Number of ways of getting 16 out of 20 correct}}{\text{Total number of ways of answering 20 questions}} = \frac{\binom{20}{16}}{2^{20}} = 0.0046
\]

**Example 1.3** Sampling without replacement. Consider an Urn Model with 10 Type I objects and 20 Type II objects, and an experiment involving sampling five objects without replacement. Let \( E = \) “precisely 2 Type I objects selected”. We need to calculate \( N \) and \( n(E) \) in order to calculate \( P(E) \). In this case \( N \) is the number of ways of choosing 5 from 30 items, and hence

\[
N = \binom{30}{5}
\]

To calculate \( n(E) \), we think of \( E \) occurring by first choosing 2 Type I objects from 10, and then choosing 3 Type II objects from 20, and hence, by the multiplication rule,

\[
n(E) = \binom{10}{2} \binom{20}{3}
\]

Therefore

\[
P(E) = \frac{\binom{10}{2} \binom{20}{3}}{\binom{30}{5}} = 0.360
\]

This result can be obtained using a conditional probability argument; consider event \( F \subseteq E \), where \( F = \) “sequence of objects 11222 obtained”. Then

\[
F = \bigcap_{i=1}^{5} F_{ij}
\]
where \( F_{ij} = \text{“type j object obtained on draw } i \text{” } i = 1, \ldots, 5, j = 1, 2 \). Then
\[
P(F) = P(F_{11})P(F_{21} | F_{11}) \cdots P(F_{52} | F_{11}, F_{21}, F_{32}) = \frac{10 \cdot 9 \cdot 20 \cdot 19 \cdot 18}{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26}
\]
Now consider event \( G \) where \( G = \text{“sequence of objects 12122 obtained”} \). Then
\[
P(G) = \frac{10 \cdot 20 \cdot 9 \cdot 19 \cdot 18}{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26}
\]
i.e. \( P(G) = P(F) \). In fact, any sequence containing two Type I and three Type II objects has this probability, and there are \( \binom{5}{2} \) such sequences. Thus, as all such sequences are mutually exclusive,
\[
P(E) = \binom{5}{2} \frac{10 \cdot 9 \cdot 20 \cdot 19 \cdot 18}{30 \cdot 29 \cdot 28 \cdot 27 \cdot 26} = \frac{\binom{10}{2} \binom{20}{3}}{\binom{30}{5}}.
\]

**Example 1.4 Sampling with replacement.** Consider an Urn Model with 10 Type I objects and 20 Type II objects, and an experiment involving sampling five objects with replacement. Let \( E = \text{“precisely 2 Type I objects selected”} \). Again, we need to calculate \( N \) and \( n(E) \) in order to calculate \( P(E) \). In this case \( N \) is the number of ways of choosing 5 from 30 items with replacement, and hence
\[
N = 30^5
\]
To calculate \( n(E) \), we think of \( E \) occurring by first choosing 2 Type I objects from 10, and 3 Type II objects from 20 in any order. Consider such sequences of selection
\[
\begin{array}{ll}
\text{Sequence} & \text{Number of ways} \\
1 \ 1 \ 2 \ 2 \ 2 & 10 \times 10 \times 20 \times 20 \times 20 \\
1 \ 2 \ 1 \ 2 \ 2 & 10 \times 20 \times 10 \times 20 \times 20 \\
\vdots & \vdots \\
\end{array}
\]
extc., and thus a sequence with 2 Type I objects and 3 Type II objects can be obtained in \( 10^2 20^3 \) ways. As before there are \( \binom{5}{2} \) such sequences, and thus
\[
P(E) = \frac{\binom{5}{2} 10^2 20^3}{30^5} = 0.329.
\]
Again, this result can be obtained using a conditional probability argument; consider event \( F \subseteq E \), where \( F = \text{“sequence of objects 11222 obtained”} \). Then
\[
P(F) = \left( \frac{10}{30} \right)^2 \left( \frac{20}{30} \right)^3
\]
as the results of the draws are independent. This result is true for any sequence containing two Type I and three Type II objects, and there are \( \binom{5}{2} \) such sequences that are mutually exclusive, so
\[
P(E) = \binom{5}{2} \left( \frac{10}{30} \right)^2 \left( \frac{20}{30} \right)^3
\]