1. BASIC CONCEPTS

1.1 Sample spaces

We will deal with a experiment which has a "random" outcome, the possible outcomes being, say, $\omega_1, \omega_2, \ldots$. The set of all possible outcomes $\Omega = \{\omega_1, \omega_2, \ldots\}$ is the **sample space** of the experiment. A particular point $\omega \in \Omega$ is an **observation**.

Examples of experiments

(i) Tossing a normal six-faced die, $\Omega = \{1, 2, \ldots, 6\}$.

(ii) Picking a card from a standard pack, Ω is the pack of 52 cards.

(iii) Administer a drug to 20 patients. For simplicity assume there are only two possible outcomes for each patient, R if the patient recovers, D otherwise. Then

$$\Omega = \{(i_1, \ldots, i_{20}); i_j = R \text{ or } D\}$$

where i_j is R or D according as patient j recovers or not, so Ω has 2^{20} points.

(iv) Pick a point from the unit interval, $[0, 1]$, then $\Omega = [0, 1]$.

All but (iv) are examples of discrete sample spaces.

For the moment Ω will be assumed to be **discrete**, i.e., it is a finite or countable set.

A subset A of Ω will be called an **event**. If the experiment is performed with outcome ω and $\omega \in A$, the event A is said to **occur**. In Example (i) A might be the event that "the outcome is even", so that $A = \{2, 4, 6\}$, in which case A occurs if one of 2,4,6 is shown on the die.

Certain set-theoretic notions have special interpretations in probability; the **complement** in Ω of the event A, A^c (alternatively A', $\overline{A}$) is the event "not A", and occurs if and only if A does not occur. The **union** $A \cup B$ of two events A and B is the event "at least one of A or B occurs". The **intersection** $A \cap B$ is the event "both A and B occur". The **inclusion** relation $A \subseteq B$ means "the occurrence of A implies the occurrence of B". Events A and B are said to be **mutually exclusive** if they are disjoint, $A \cap B = \emptyset$, and so both cannot occur together.

1.2 Classical probability

Classical probability was concerned with modelling situations where there are just a finite number of possible outcomes of the experiment and each of these outcomes is "equally likely". For example, tossing a fair coin or an unloaded die, or picking a card from a standard well-shuffled pack. Here Ω is a finite set with N points, say, $\Omega = \{\omega_1, \ldots, \omega_N\}$, and attached to each event $A \subseteq \Omega$ is a measure of how "likely" the event A is to occur. Here we take $\mathbb{P}(A)$, the **probability** of the event A to be

$$\mathbb{P}(A) = \frac{N_A}{N} = \frac{\text{number of points in } A}{\text{number of points in } \Omega}$$

where N_A is the number of points in A. Thus the probability $\mathbb{P}(\cdot)$ is just a real-valued function defined on subsets A of Ω.

Properties of $\mathbb{P}(\cdot)$

1. $0 \leqslant \mathbb{P}(A) \leqslant 1$.
2. $\mathbb{P}(\Omega) = 1$.
3. If $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

In a while we will take properties 1, 2 and an extension of 3 as the basis of an axiomatic development of probability in a more general context. The following properties follow immediately from the definition of $\mathbb{P}(\cdot)$ above, but they may be deduced easily from 1, 2 and 3 without reference to the definition.

4. $\mathbb{P}(\emptyset) = 0$.
5. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
6. If $A \subseteq B$ then $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.
7. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
8. $\mathbb{P}(A \cup B) \leqslant \mathbb{P}(A) + \mathbb{P}(B)$.

1.3 Combinatorial analysis

Fundamental rule: There are r multiple choices to be made in sequence: there are m_1 possibilities for the first choice; after making the first choice there are m_2 possibilities for the second choice; after making the first two choices there are m_3 possibilities for the third choice, and so on until after making the first $r-1$ choices there are m_r possibilities for the rth choice. Then the total number of different possibilities for the set of choices is

$$m_1 m_2 \cdots m_r$$

Example 1.1 A restaurant menu has 6 starters, 7 main courses and 6 puddings. The total number of different three-course meals that may be served is $6 \times 7 \times 6 = 252$. □

Sampling models

Many of the standard calculations that arise in classical probability where outcomes need to be counted may be put in a standard framework of **sampling**. Think of drawing m balls from an urn which initially contains n distinguishable balls (they are numbered 1 to n, say). This may be done in a number of ways:

1. Sampling with replacement and with ordering. The balls are replaced between successive draws and the order in which balls are drawn is noted. Then the Fundamental Rule shows that there are n^m possible ways.

2. Sampling without replacement and with ordering $(m \leqslant n)$**.** The balls are not replaced after drawing and the order is noted. Then the number of ways is

$$n\,(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!} = \,{}^nP_m;$$

the symbol ${}^nP_m$ represents the number of permutations of n objects m at a time.

An important special case occurs when $n = m$. We get the number of permutations of n distinguishable objects (that is the number of distinguishable ways they may be laid out in a line, say) is $n!$.

3. Sampling without replacement and without ordering. If we take m balls from n and they were ordered there would be $n(n-1)\cdots(n-m+1) = n!/(n-m)!$ ways but each unordered selection may be permuted in $m!$ different ways (or give $m!$ ordered arrangements) so that the total number of unordered ways is

$$\frac{n!}{(n-m)!\,m!} = \binom{n}{m} = \,{}^nC_m;$$

the symbol $\binom{n}{m}$ is the binomial coefficient, usually read "n choose m", and represents the number of ways of picking m objects from n; note that $\binom{n}{m}$ is the coefficient of x^m in the expansion of $(1+x)^n$ using the binomial theorem.

Suppose that, of the n balls, m_1 are of colour 1, m_2 are of colour 2 and so on up to m_k balls of colour k, where $n = m_1 + \cdots + m_k$. Consider the number of permutations of the n balls when the balls are distinguishable only by colour. For example, if $n = 4$ and there are two black balls and two white balls then there are 6 possible arrangements:

$$BBWW \quad BWBW \quad BWWB \quad WBBW \quad WBWB \quad WWBB$$

If the balls are distinguishable then there are $n!$ permutations, but within colour i there are $m_i!$ ways of permuting the balls so the Fundamental Rule gives that for each distinguishable arrangement there are $\prod_1^k (m_i!)$ ways of permuting the balls leaving the arrangement the same if the balls are distinguishable only by colour. Thus the number of arrangements distinguishable only by colour is

$$\frac{n!}{m_1!\, m_2! \cdots m_k!} = \binom{n}{m_1 \cdots m_k},$$

which is the multinomial coefficient (the coefficient of $x_1^{m_1} x_2^{m_2} \cdots x_k^{m_k}$ in the expansion of $(x_1 + \cdots + x_k)^n$). An alternative way of seeing this is to think of first choosing the positions for the balls of colour 1, which can be done in $\binom{n}{m_1}$ ways, then choose the positions for the balls of colour 2 in $\binom{n-m_1}{m_2}$ ways and so on to see that the total number of ways, using the Fundamental Rule, is

$$\binom{n}{m_1}\binom{n-m_1}{m_2}\binom{n-m_1-m_2}{m_3}\cdots\binom{n-m_1-\cdots-m_{k-1}}{m_k} = \binom{n}{m_1 \cdots m_k}.$$

4. Sampling with replacement and without ordering. Draw m balls one after another, each time noting the number of the ball and replacing it before the next draw,

Ball No.	1	2	3	4	$\cdots$	n
Times drawn	✓	✓✓	✓	✓✓	$\cdots$	✓

so that the number of ✓ is the number of times the ball is drawn. The number of ways is then the number of ways that $n-1$ vertical lines may be put between m ✓ which is

$$\binom{n+m-1}{n-1};$$

think of choosing $n-1$ slots for the vertical lines out of $n+m-1$ slots and then the checks go in the remainder.

Allocation models

An alternative way of thinking of these counting schemes is to think about allocating m tokens (labelled $1,\ldots,m$) to n boxes (labelled $1,\ldots,n$). The four cases considered previously correspond to:

1. Each box may contain any number of tokens and the labels on the tokens are observed.

2. No box may contain more than one token and the labels on the tokens are observed.

3. No box may contain more than one token and the labels on the the tokens are **not** observed.

4. Each box may contain any number of tokens and the labels on the tokens are **not** observed.

In the next examples in which each outcome in the probability space is equally likely we will calculate the probability of the event A by computing the number of points N in the sample space Ω and then computing the number N_A of points in A. In each case the probability will then be

$$\mathbb{P}(A) = \frac{N_A}{N} \,.$$

Example 1.2 What is the probability that a Poker hand shows five different face values? (A poker hand contains 5 cards.) One way to do this is to think of Ω consisting of all possible poker hands, that is, unordered sets of five cards chosen from a standard pack of 52. We have $N = \binom{52}{5}$ and A consists of those hands showing five different face values so that

$$N_A = \binom{13}{5} \times 4^5$$

since we may think of first choosing the unordered sets of 5 face values and then the different suits for each value, and use the Fundamental Rule. An alternative way in this problem, and in other situations involving sampling without replacement, is to think of Ω comprising all ordered sets of 5 cards (imagine the cards being dealt in sequence). Then $N = 52 \times 51 \times 50 \times 49 \times 48$ and $N_A = 52 \times 48 \times 44 \times 40 \times 36$, thinking of the possible choices for the first card, second card in the hand and so on to give 5 different face values. You should check that these two different approaches give the same probabilities. Either approach is fine, but remember to be consistent – if your sample space has unordered (respectively, ordered) points then the points in A must be unordered (respectively, ordered). □

Example 1.3 What is the probability that a Bridge hand (13 cards) contains 5 spades ($\spadesuit$), 3 hearts ($\heartsuit$), 4 diamonds ($\diamondsuit$) and 1 club ($\clubsuit$)? Take Ω to be the set of unordered Bridge hands, so we have $N = \binom{52}{13}$ while

$$N_A = \binom{13}{5} \times \binom{13}{3} \times \binom{13}{4} \times \binom{13}{1}$$

since we may think of choosing in sequence the spades, hearts, diamonds and club and then use the Fundamental Rule to get the total number of ways. □

Example 1.4 If there are r people in a room, what is the probability that at least two have the same birthday? In this case the number of points in the sample space Ω is $N = (365)^r$, since the sample space consists of all possible r-tuples of birthdays. If A is the required event, this is a case where it is easier to calculate the number of points in A^c, the complement of A, since A^c is just the event that no two of the r people share a birthday. We then have

$$N_{A^c} = 365 \times 364 \times 363 \times \cdots \times (365 - r + 1),$$

and so

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{365 \times 364 \times 363 \times \cdots \times (365 - r + 1)}{(365)^r} = p_r, \quad \text{say.}$$

It is interesting to note the following values

r	10	15	20	21	22	23	24	25	30	40	55
p_r	0·12	0·25	0·41	0·44	0·48	0·51	0·54	0·57	0·71	0·89	0·99

which shows the well known fact that if there are 23 or more people in a room there is higher than evens chances that at least 2 share a birthday. □

Theorem 1.5 *(Stirling's Formula) As $n \to \infty$,*

$$\log\left(\frac{n!\,e^n}{n^{n+\frac{1}{2}}}\right) = \log\left(\sqrt{2\pi}\right) + O\left(1/n\right).$$

Proof. Let $c_n = \log(n!) + n - \left(n + \frac{1}{2}\right)\log n$, then

$$c_n - c_{n+1} = (n + \tfrac{1}{2})\log(1 + 1/n) - 1. \tag{1.6}$$

For $0 < x < 1$, if we subtract the two expressions

$$\log(1+x) - x = -\frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$
$$\log(1-x) + x = -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots$$

and divide by $2x$, we obtain

$$\frac{1}{2x}\log\left(\frac{1+x}{1-x}\right) - 1 = \frac{x^2}{3} + \frac{x^4}{5} + \frac{x^6}{7} + \cdots \tag{1.7}$$
$$\leqslant \frac{x^2}{3} + \frac{x^4}{3} + \frac{x^6}{3} + \cdots = \frac{1}{3}\frac{x^2}{1-x^2}. \tag{1.8}$$

Now put $x = 1/(2n+1)$ and we see from (1.6) that the left-hand side of (1.7) is then $c_n - c_{n+1}$, and furthermore from (1.7) that $c_n - c_{n+1} \geqslant 0$ and from (1.8) that

$$c_n - c_{n+1} \leqslant \frac{1}{12}\left(\frac{1}{n} - \frac{1}{n+1}\right).$$

This shows that c_n is monotone non-increasing in n and $c_n - 1/(12n)$ is monotone non-decreasing so that c_n is bounded below and hence converges $c_n \to c$, for some c.

To determine the value of c, define $I_r = \int\limits_0^{\pi/2} \sin^r \theta d\theta$, for $r \geqslant 0$, so that $I_0 = \pi/2$ and $I_1 = 1$. Integrating by parts for $r > 1$, we have

$$I_r = \left[-\sin^{r-1}\theta\cos\theta\right]_0^{\pi/2} + (r-1)\int_0^{\pi/2} \sin^{r-2}\theta\cos^2\theta d\theta = (r-1)\left(I_{r-2} - I_r\right),$$

so that $rI_r = (r-1)I_{r-2}$. It is immediate that $I_{2n+1} \leqslant I_{2n} \leqslant I_{2n-1}$, hence

$$1 \leqslant \frac{I_{2n}}{I_{2n+1}} \leqslant \frac{I_{2n-1}}{I_{2n+1}} = \frac{2n+1}{2n} \to 1, \quad \text{as} \quad n \to \infty.$$

Calculate $$I_{2n} = \frac{2n-1}{2n} I_{2n-2} = \frac{(2n-1)(2n-3)\cdots 1}{(2n)(2n-2)\cdots 2} I_0 = \frac{(2n)!}{(2^n n!)^2}\frac{\pi}{2}$$

and $$I_{2n+1} = \frac{2n}{2n+1} I_{2n-1} = \frac{(2n)(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 1} I_1 = \frac{(2^n n!)^2}{(2n+1)!}.$$

Dividing these we see that, as $n \to \infty$,

$$\frac{I_{2n}}{I_{2n+1}} = \frac{(2n+1)((2n)!)^2}{(2^n n!)^4}\frac{\pi}{2} \to 1, \quad \text{or} \quad \frac{(2^n n!)^2}{(2n)!}\frac{1}{\sqrt{n}} \to \sqrt{\pi}.$$

Now note that

$$2c_n - c_{2n} = \log\sqrt{2} + \log\left(\frac{(2^n n!)^2}{(2n)!}\frac{1}{\sqrt{n}}\right) \to \log\sqrt{2} + \log\sqrt{\pi} = \log\sqrt{2\pi},$$

so that $c = \lim\limits_{n\to\infty}(2c_n - c_{2n}) = \log\sqrt{2\pi}$, as required. □

The most common statement of Stirling's formula is given as a corollary.

Corollary 1.9 *As $n \to \infty$, $n! \sim \sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}}$, where, in this context, $\sim$ indicates that the ratio of the two sides tends to 1.*

Example 1.10 Suppose that $4n$ balls, of which $2n$ are red and $2n$ are black, are put at random into two urns, so that each urn contains $2n$ balls. What is the probability that each urn contains n red balls and n black balls? The probability is

$$\begin{aligned} p_n &= \binom{2n}{n}\binom{2n}{n}\Big/\binom{4n}{2n} = \left(\frac{(2n)!}{n!}\right)^4 \frac{1}{(4n)!} \\ &\sim \left(\frac{\sqrt{2\pi}e^{-2n}(2n)^{2n+\frac{1}{2}}}{\sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}}}\right)^4 \frac{1}{\sqrt{2\pi}e^{-4n}(4n)^{4n+\frac{1}{2}}} = \sqrt{\frac{2}{\pi n}} = a_n, \text{ say.} \end{aligned}$$

We may calculate the exact probability p_n, its approximant a_n from Stirling's formula and their ratio for different values of n.

n	1	10	13	20	40	60	100
p_n	0·667	0·248	0·218	0·177	0·126	0·103	0·080
a_n	0·798	0·252	0·221	0·178	0·126	0·103	0·080
p_n/a_n	0·836	0·981	0·986	0·991	0·995	0·997	0·998

The case $n = 13$ corresponds to dividing a standard pack of cards in two and obtaining equal numbers of red and black cards in each half. □