SECTION 1 : Mathematical Building Blocks

1.1 : Probability

The theory of probability is a branch of *pure* mathematics. This means that we could deduce laws and theorems which describe how to manipulate probabilities, starting from a set of axioms – in a manner similar to the theory of arithmetic. Many modern textbooks take this approach, since it allows the mathematical machinery of **measure theory** to be directly applied.

The disadvantage to such an approach is that it is both abstract and complicated. We will instead try to develop ideas about probability which are more **intuitive**. (c.f. how we learn to count at school).

1.1.1 : Counting (combinatorial) Definition of Probability

Suppose we observe some event (e.g. a physical experiment) for which there are a finite number, n, of possible outcomes. Suppose the outcomes can be grouped together according to some well-defined attribute or characteristic. e.g.:-

Event	example attributes
Tossing a coin	head, tail
Throwing a dice	1, 2, 3, odd number, even number

Suppose that attribute A occurs in m of the n possible outcomes. Then we could define the **probability** of an outcome having attribute A – which we write simply as $P(A)$ – as:-

$$P(A) \quad = \quad \frac{\text{number of outcomes with attribute } A}{\text{total number of outcomes}} \quad = \frac{m}{n}$$

For example, if a coin is equally likely to fall as a head or tail, we say that:-

$$P(\text{head}) = P(\text{tail}) = 1/2$$

1.1.2 : 'Frequentist' Definition of Probability

How do we 'know' from the outset that a coin is equally likely to fall as a head or a tail? In truth we do *not* know this *a priori*, but our intuition might lead us to reason as follows. Suppose we toss the coin a large number of times and, in the long run, the coin falls as a head half of the time and as a tail half of the time; we could then regard a head and a tail as equally probable outcomes. This intuitive idea of what happens to the coin when it is tossed a large number of times forms the basis of what is known as the **frequentist** definition of probability.

More generally, suppose we perform an experiment N times. (This can be something as simple as tossing a coin, or something as complex as measuring the Hubble constant). We define the **relative frequency** of an outcome with attribute A_i as:-

$$\text{rel. freq.}(A_i) \quad = \quad \frac{\text{number of outcomes with attribute } A_i}{\text{total number of outcomes}} \quad = \quad \frac{n(A_i)}{N}$$

We then define the probability of outcome A_i as

$$P(A_i) \quad = \quad \lim \frac{n(A_i)}{N} \qquad \text{as} \quad N \to \infty$$

Aside: Later in the course we will consider how to *test* e.g. whether a coin is **fair** (i.e. $P(\text{head}) = P(\text{tail}) = 1/2$) by asking how close the experimentally determined ratio, $n(\text{head})/N$, should be to $1/2$, for a given number of 'experiments', N, in order to be confident that the coin is fair. We will see that we can never be **absolutely** sure that the coin is fair, but statistics allows us to make quantitative statements about how **likely** it is that the coin is fair. Here the assumption of a fair coin ais an example of a **hypothesis**, which we can test by tossing the coin a large number of times. Based on our accumulated data, we then decide to either accept or reject the hypothesis of a fair coin.

Tossing a coin is an example of a simple event because there are only two possible outcomes and these are mutually exclusive. Generally, however, we must deal with events which are *not* simple, but rather are *composite* – i.e. combinations of two or more simple events. Consider, for example, a pack of cards. (Assume that the probability of drawing each card is $1/52$). Three such composite events would be:-

- Probability of drawing an ace **or** a spade
- Probability of drawing an ace **and** a spade
- Probability of drawing an ace **then** a spade

To handle such events we need laws for combining probabilities. We will not *prove* these, but justify them by counting arguments – essentially using our well-known ideas about the intersection and union of sets.

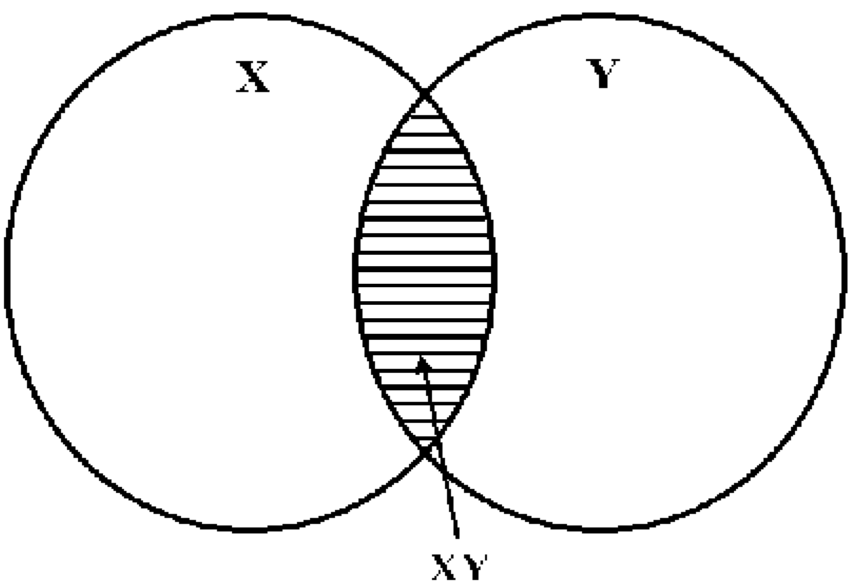
1.1.3 : Law of Addition

Let X and Y be two different sets of outcomes of an experiment. Let $X+Y$ denote the set of outcomes which occur in *either* X *or* in Y, and XY the set of outcomes which occur in *both* X *and* Y. Then

$$P(X+Y) \quad = \quad P(X) \; + \; P(Y) \; - \; P(XY)$$

We can justify this equation by counting arguments. Suppose we carry out the experiment N times. let $n(X)$, $n(Y)$, $n(X+Y)$ and $n(XY)$ denote the number of elements in the sets X, Y, $X+Y$ and XY respectively (c.f. Figure 1).

Figure 1: Venn diagram showing two intersecting sets of outcomes



Simple counting gives

$$n(X+Y) \; = \; n(X) \; + \; n(Y) \; - \; n(XY)$$

Dividing by N and letting $N \to \infty$, we obtain the law of addition. Thus, in order to determine the probability that an outcome belongs to set X *or* set Y, we add the probability that the outcome belongs to set X to the probability that it belongs to set Y. But this means that we have counted *twice* those outcomes which belongs to both X and Y, so we need to *subtract*, $P(XY)$.

Ex: $P(\text{ace } \textbf{or} \text{ spade}) = P(\text{ace}) + P(\text{spade}) - P(\text{ace } \textbf{and} \text{ spade})$
$= 4/52 + 13/52 - 1/52$
$= 16/52 = 4/13$

1.1.4 : Conditional Probability

Consider an experiment which is repeated n times – i.e. we have a total of n outcomes. Let n_1 of these outcomes have some attribute A_1, n_2 have another attribute A_2 and n_{12} have attributes A_1 *and* A_2. Then,

$$P(A_1) = \frac{n_1}{n} \quad (\text{strictly } P(A_1) = \lim \frac{n_1}{n}, \text{ as } N \to \infty)$$

Also

$$P(A_2) = \frac{n_2}{n}$$

and

$$P(A_1 \text{ and } A_2) = \frac{n_{12}}{n}$$

We can write this last equation as

$$P(A_1 \text{ and } A_2) = \frac{n_{12}}{n_1}\frac{n_1}{n} = \frac{n_{12}}{n_1} P(A_1)$$

n_{12}/n_1 is the relative frequency of those outcomes which have attribute A_1, which **also** have attribute A_2.

In the limit as $n_1 \to \infty$, n_{12}/n_1 is defined as the **conditional probability** of the outcome having attribute A_2, *given* that it has attribute A_1. It is usually written as $P(A_2|A_1)$.

1.1.5 : Law of Multiplication

In the above notation

$$P(A_1 \text{ and } A_2) = P(A_1\, A_2) = P(A_1)\, P(A_2|A_1) = P(A_2)\, P(A_1|A_2)$$

Thus

$$P(A_2|A_1) = \frac{P(A_1\, A_2)}{P(A_1)}$$

which is often how conditional probabilities are defined in practice.