PROBABILITY THEORY

In the world of Statistics, the word that can most suitably replace *probability* would be ' chance '. Probability can be considered as a measure of likelihood of an event occurring. There are two ways of assigning numerical values to probability : the *classical* approach and *empirical* method.

Before probing into the core of this topic, it will be most appropriate to, first of all, make the reader familiar with the different notations that will be used throughout it. It must be mentioned, here, that *probability theory* is quite closely linked to *set theory*. It has, therefore, been considered necessary to remind the reader of the equivalence of the terms of these two concepts.

Notation

S - sample space (universal set)
$A, B, C, \ldots$ - events defined on the sample space S (subsets of the universal set)
$n(A)$ - number of outcomes favourable to A (number of elements in set A)
$p(A)$ - probability of event A occurring
f - null event (empty set)
$\overline{A}$ or A' - A complement, that is, the event which automatically occurs in the absence of event A (the set of all elements not belonging to set A)
$\cup$ - union
$\cap$ - intersection
$\subset$ - ' is a subset of '
$\subseteq$ - ' is a subset of ' or ' is equal to '

Classical method of obtaining probability

Given an event A defined on a sample space S, the probability of A occurring, denoted by $p(A)$, is defined as

$$p(A) = \frac{n(A)}{n(S)}\,.$$

It should be clear that $n(S)$ represents the total number of outcomes.

In fact, the classical definition of probability applies when the outcomes of an experiment are equally likely to occur. When deriving mathematical rules for probability, it is useful to consider the classical definition of probability.

Example

A six-sided unbiased (ordinary fair) die is tossed. What is the probability of obtaining a multiple of 3 ?

Solution

The sample space $S = \{1,2,3,4,5,6\} \Rightarrow n(S) = 6$;
Let A = ' multiple of 3 '
Therefore, $A = \{3,6\} \Rightarrow n(A) = 2$;
It follows that $p(A) = \dfrac{n(A)}{n(S)} = \dfrac{2}{6} = \dfrac{1}{3}$.

Empirical method of obtaining probability

This definition of probability lays emphasis on a much more practical rather than theoretical approach. Here, probability is determined as a result of a large number of repeated experiments under presumably *identical* conditions.

For example, let us try to find out whether a coin is unbiased. The most natural thing to do is to toss it a certain number of times n and record the number of 'heads', $n(H)$, or number of 'tails' , $n(T)$). A ratio $n(H) : n(T)$ approximately equal to 1 : 1 would suggest that the coin is *probably* unbiased. On the other hand, if the coin is tossed 20 times, for example, and 15 heads are recorded, it cannot be immediately concluded that the coin is biased. This is because the value of n is too small to allow any hasty decision. It is expected that, as we increase the number of tosses, the value of $\dfrac{n(H)}{n}$ will tend to the real value of $p(H)$ for the coin.

Let us consider the following table, which gives the results of tosses of a coin generated at random by a computer.

n	10	50	100	500	1000	5000
$n(H)$	7	33	62	289	572	2797
$p(H)$	0.7	0.66	0.62	0.578	0.572	0.5594

n	10000	50000	100000	500000	1000000	5000000
$n(H)$	5521	27530	54997	274895	550020	2750050
$p(H)$	0.5521	0.5506	0.54997	0.54979	0.55002	0.55001

It can be noted that $p(H)$ approaches 0.55 as the number of tosses increases indefinitely. We can, within some security margin, conclude that the coin is slightly biased towards 'heads'.

Subjective method of obtaining probability

There are cases where all outcomes are not equally likely to occur and also where a good estimate of probability cannot be obtained because an experiment cannot be repeated under identical conditions. An example would be to calculate the probability that a shop will sell exactly 10 television sets on a particular day. In those cases, we are forced to form a subjective probability, based on past experience, records, expert opinion or other factors. This method obviously has a very large margin of error but it is sometimes the only method available.

AXIOMS OF PROBABILITY

1. For any sample space S, $p(S) = 1$.
2. For any given event A defined on a sample space S, $0 \leq p(A) \leq 1$.
3. For any two *mutually exclusive* events A and B defined on a sample space S,
$$p(A \cup B) = p(A) + p(B)$$

Note

Two events are said to be mutually exclusive if they have no intersection.

Proofs

1. $p(S) = 1$.

A probability of 1 is known as *certainty*. It is obvious that the outcome of an experiment must belong to the sample space corresponding to that particular experiment. Hence, the sample space occurs all the time with 100% probability. If not, it would not have been well-defined !

2. $0 \leq p(A) \leq 1$

This proof consists of two parts :

(i) $p(A) \geq 0$

From the classical definition, $p(A) = \dfrac{n(A)}{n(S)}$. Since $n(A)$ and $n(S)$ are both *natural* numbers, it goes without saying that $p(A) \geq 0$.

(ii) $p(A) \le 1$

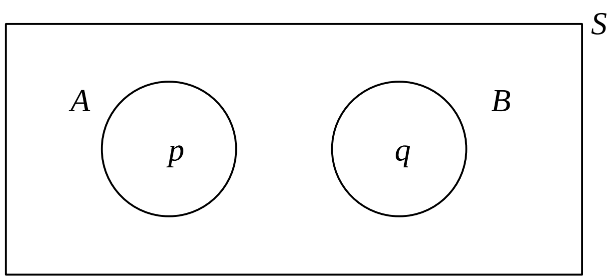
Here, we know that $A \subseteq S$. Therefore, $n(A) \le n(S)$. Dividing both sides of the inequality by $n(S)$, we have the required result :

$$n(A) \le n(S)$$
$$\Rightarrow \frac{n(A)}{n(S)} \le \frac{n(S)}{n(S)}$$
$$\Rightarrow p(A) \le 1.$$

We conclude that probability can only take on values between 0 and 1 inclusive.

3. $p(A \cup B) = p(A) + p(B)$

This axiom can easily be proved by using a Venn diagram.



Let $n(A) = p$ and $n(B) = q$.
Thus, $n(A) + n(B) = p + q$ and also, $n(A \cup B) = p + q$.

We can therefore write $n(A \cup B) = n(A) + n(B)$ and, dividing both sides by $n(S)$, we have

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)}$$
$$\Rightarrow p(A \cup B) = p(A) + p(B).$$

Now that we have proved the axioms, let us have a look at some further rules of probability. Most of them are derived from the axioms themselves.

Further rules of probability

1. $p(f) = 0$.

There are two ways of proving the above :

(i) $p(f) = \dfrac{n(f)}{n(S)} = \dfrac{0}{n(S)} = 0$.

(ii) Since S and f are complementary, they are also mutually exclusive. Using the third axiom of probability, we have

$$p(S \cup f) = p(S) + p(f)$$

Since $S \cup f = S$, it follows that

$$p(S) = p(S) + p(f)$$
$$\Rightarrow p(f) = p(S) - p(S) = 0$$

2. $p(A') = 1 - p(A)$

We know that A and A' are mutually exclusive. By the third axiom,

$$p(A \cup A') = p(A) + p(A')$$

Since $A \cup A' = S$,

$$p(S) = p(A) + p(A')$$
$$\Rightarrow 1 = p(A) + p(A') \text{ from the first axiom and}$$
$$p(A') = 1 - p(A).$$

3. $p(A) = p(A \cap B) + p(A \cap B')$
(It can easily be checked that $n(A) = n(A \cap B) + n(A \cap B')$).

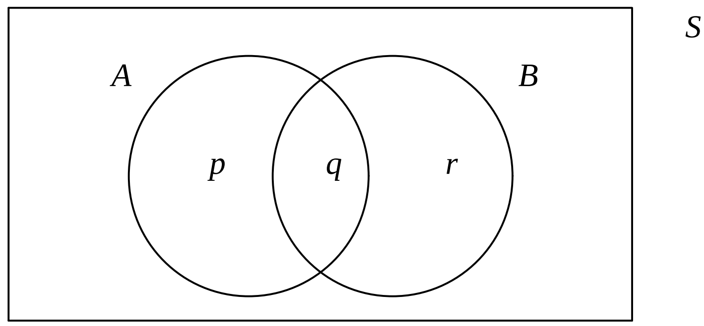
4. De Morgan's rules.

(i) $p(A' \cap B') = p(A \cup B)'$
(ii) $p(A' \cup B') = p(A \cap B)'$

5. General addition rule for two events.

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

This rule will also be proven by the use of a Venn diagram. From the diagram below, $n(A) = p + q$, $n(B) = q + r$ and $n(A \cap B) = q$.
We have that $n(A) + n(B) - n(A \cap B) = (p + q) + (q + r) - q = p + q + r$.
Also, $n(A \cup B) = p + q + r$.



Therefore, $n(A) + n(B) - n(A \cap B) = n(A \cup B)$. Dividing by $n(S)$ on both sides, we obtain

$$p(A) + p(B) - p(A \cap B) = p(A \cup B).$$

CONDITIONAL PROBABILITY

Conditional probability theory helps to calculate the probability of an event occurring given that another event has occurred. In a sense, it verifies whether an event *A* is independent of a second event *B*. We use conditional probability very often in solving basic statistical problems; if asked, for example, the probability of drawing two red balls from a bag containing 3 blue and 5 red balls, we would find it very straightforward to reply $\frac{5}{8}\times\frac{4}{7}=\frac{5}{14}$. However, there is a rich theoretical background behind this simple calculation, as will be seen later.

Let us start with a real-life situation where the event *A* is defined as "Mauritius will produce 600 000 tonnes of sugar in 2004" and *B* is the event that "there will be a heavy cyclone during the year 2004". It is obvious that both *A* and *B* have their individual probabilities of occurring but it should also be clear that the probability of *A* will change depending on whether *B* occurs, that is, a cyclone could definitely bring about a drastic decrease in the number of tonnes of sugar produced. This is a case of conditional probability.

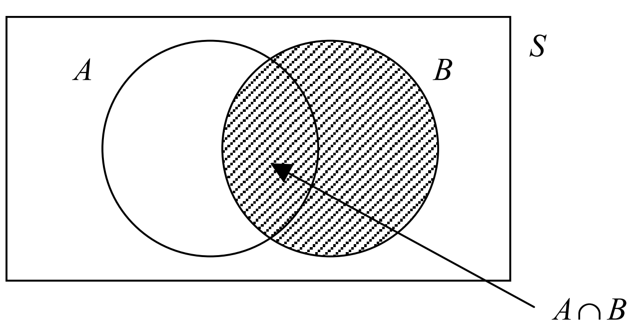
Notation

$P(A \mid B)$ means "the probability that *A* occurs given that *B* has already occurred"
[Remember that we are only calculating the probability that *A* occurs (nothing to do with *B*).]

Definition

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

The above formula can be explained by means of the following Venn diagram.



If event B occurs, obviously its complement B' cannot occur any more. Thus, if any subsequent event has to occur, it can only do so in B. Therefore, the initial sample space S has now been reduced to the new sample space B. Now, what is the probability of A occurring in B? The answer is clearly the part of A which also belongs to B, that is, $A \cap B$ (indicated by an arrow in the above diagram).

Hence, $P(A \mid B) = \dfrac{n(A \cap B)}{n(B)} = \dfrac{n(A \cap B) \div n(S)}{n(B) \div n(S)} = \dfrac{P(A \cap B)}{P(B)}$.

It is worth noting that, sometimes, we do not have to solve a problem by strictly using theory so that a logical approach may also lead us to the solution more easily. Example 1 illustrates both approaches to the same problem.

Example 1

An ordinary fair (six-sided unbiased) die is tossed. If the score is even, what is the probability that it is also prime?

Solution

Method 1 (Theory)

Let A = "the score is even" and
B = "the score is prime"

We wish to calculate $P(B \mid A)$, that is, $\dfrac{P(B \cap A)}{P(A)}$.

The sample space of scores is $S = \{1, 2, 3, 4, 5, 6\}$.
Note that $A = \{2, 4, 6\}$, $B = \{2, 3, 5\}$ and $A \cap B = \{2\}$.
$P(A)$, the probability that the score is even, is $\frac{3}{6}$.

$P(A \cap B)$, the probability that the score is both even and prime, is $\frac{1}{6}$.

Therefore, $\dfrac{P(B \cap A)}{P(A)} = \dfrac{\left(\frac{1}{6}\right)}{\left(\frac{3}{6}\right)} = \frac{1}{3}$.

Method 2 (Logic)

We know that the score is even so that any subsequent outcome must belong to the set {2, 4, 6}, which is the *new sample space* (read discussion above).

The only prime number in this set is 2 and the probability that it occurs is clearly $\frac{1}{3}$.

(Short and sweet!)

A problem on conditional probability can also be solved by means of a *tree diagram* as shown in Example 2 below.

Example 2

A man goes to work on foot, by bus or by car with respective probabilities of 0.5, 0.2 and 0.3 respectively. If he goes on foot, the probability that he arrives to work late is 0.4. If he goes by bus, the probability that he arrives to work late is 0.7 and if he goes by car, the probability that he arrives to work late is 0.5. Determine the probability that

(a) he is late on a given day,

(b) he travelled by bus given that he is late.

[To avoid any confusion, you may assume that being exactly on time is the same as being early.]

Solution

Let F = "man goes on foot"

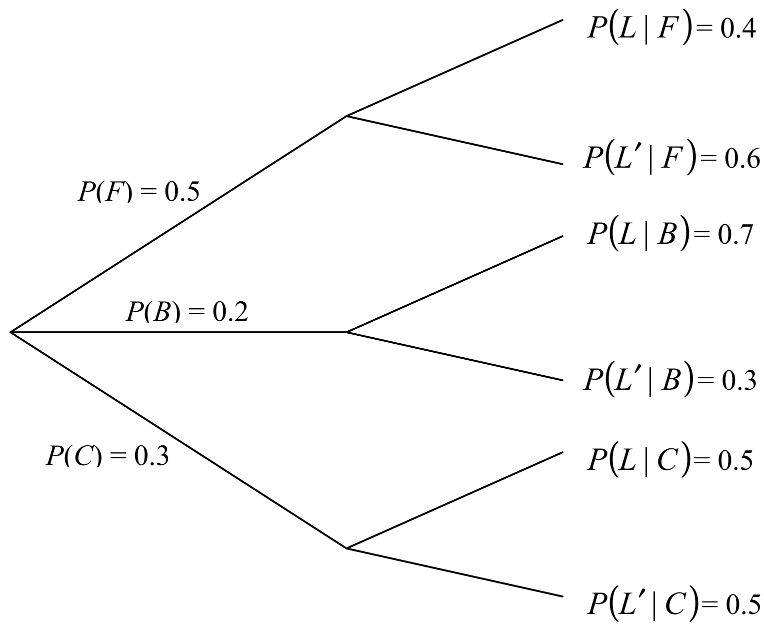
B = "man travels by bus"

C = "man travels by car"

L = "man is late"

Note that there is no need to define an event E, for example, where E = "man is early" since that event is simply the complement of "man is late", hence, denoted by L'.

The information can be illustrated by a tree diagram.



(a) The probability of the man being late, *irrespective of the means of transport used*, is given by

$$P(L) = P(F)\,P(L \mid F) + P(B)\,P(L \mid B) + P(C)\,P(L \mid C)$$
$$= (0.5)(0.4) + (0.2)(0.7) + (0.3)(0.5) = 0.49$$

(b) $$P(B \mid L) = \frac{P(B \cap L)}{P(L)} = \frac{P(B)\,P(L \mid B)}{P(F)\,P(L \mid F) + P(B)\,P(L \mid B) + P(C)\,P(L \mid C)}$$

$= \dfrac{0.14}{0.49} = \dfrac{2}{7}$. (this could mean the contribution of bus in lateness)

The next example shows the application of probability in the case of a *contingency table*. A contingency table is just a table of frequencies representing two factors in terms of their attributes.

Example 3

The following table shows the frequency distribution of grades obtained in Mathematics by students of different sections of a certain form in a secondary school.

	A	*B*	*C*
Form V Red	15	25	40
Form V Blue	26	44	10

If a student is chosen at random, find the probability that the student

(a) obtained an *A*
(b) is from Form V Blue
(c) is from Form V Red and obtained a *C*
(d) obtained an *A* given that he is from Form V Red
(e) is from Form V Blue given that he obtained a *B*.

Solution

We start by calculating the *marginal* and *grand* totals.

	A	*B*	*C*	**Total**
Form V Red	15	25	40	**80**
Form V Blue	26	44	10	**80**
Total	**41**	**69**	**50**	**160**

(a) P [student obtained an A] = $\frac{41}{160}$.

(b) P [student is from Form V Blue] = $\frac{80}{160} = \frac{1}{2}$.

(c) P [student is from Form V Red and obtained a C] = $\frac{40}{160} = \frac{1}{4}$.

(d) P [student obtained an A given that he is from Form V Red] = $\frac{15}{80} = \frac{3}{16}$.

(e) P [student is from Form V Blue given that he obtained a B] = $\frac{44}{69}$.

Independence

Imagine that an event A is independent of another event B. It can be easily understood that the probability of A occurring will be unaffected by the fact that B has occurred or not. Thus, we simply conclude that

$$P(A \mid B) = P(A).$$

From definition, we obtain

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

so that $P(A \cap B) = P(A)\ P(B)$.

The above result is known as the **multiplicative rule** for two independent events. Note that *independent* events are **not** *mutually exclusive* events! The difference is in fact very obvious: for mutually exclusive events, there is no intersection whereas independent events definitely do have an intersection (the above result says it all!)