PROBABILITY

1 INTRODUCTION

Some events seem to have a certainty about their outcome; while a few are certain not to happen. There are others, which, with regard to their outcome, vary between the two extreme situations referred to above. In a rough way, a measure of the extent of the happening or non-happening of an event may be said to be given by the term **Probability.** The word probability and the word chance are synonymous and may be taken, in this context, to be indistinguishable.

Probability, in the conventional sense, is a ratio between what may be called as the 'number of favorable' to the 'total number'; and, as such this ratio is a positive fraction varying in value between zero, when the events is certain not to happen, to one when an event is certain to happen.

Thus the probability of having a white ball from a bag containing all black balls is zero, while that of drawing a black ball is unity.

2 CONCEPT OF PROBABILITY IN SET THEORETIC LANGUAGE

2.1 RANDOM EXPERIMENT

It is an operation which can result in any one of its well defined outcomes and the outcome cannot be predicted with certainty.

For example some of the random experiments are

- toss of a coin, which can result in either a head or a tail
- throw of a die which can result in any one of the six faces
- drawing a card from a pack of 52 cards which can result in any one of the 52 cards.

2.2 SAMPLE SPACE AND SAMPLE POINTS

The set of all possible outcomes of a random experiment is called sample space and is denoted by S. Every possible outcome i.e. every element of this set is a sample point.

For example:

- In toss of a coin, $S = \{H, T\}$ where H and T are sample points representing a head and a tail respectively.
- In throw of a die, $S = \{1, 2, 3, 4, 5, 6\}$ where the numbers are the sample points representing the six faces.

2.3 TRIAL

When an experiment is repeated under similar conditions and it does not give the same result each time but may result in any one of the several possible outcomes, the experiment is called a trial and the outcomes are called cases. The number of times the experiment is repeated is called the number of trials.

For example:

- One toss of coin is a trial when coin is tossed 5 times.
- One throw of a die is a trial when the die is thrown 4 times.

2.4 EVENT

A subset of sample space, i.e. a set of some of possible outcomes of a random experiment is called as event.

Simple event

Each sample point in the sample space is called an elementary event or simple event. For example occurrence of head in throw of a coin is simple event.

Sure event

The set containing all sample points is a sure event as in the throw of a die the occurrence of natural number less than 7, is a sure event.

Null event

The set which does not contain any sample point.

Mixed/compound event

A subset of sample space S containing more than one element is called a mixed event or a compound event.

Compliment of an event

Let S be the sample space and E be an event then E^c or $\overline{E}$ represents complement of event E which is a subset containing all sample points in S which are not in E. It refers to the non occurrence of event E.

2.5 ALGEBRA OF EVENTS

In connection with basic probability laws we shall need the following concepts and facts about events (subsets) A, B, C, …… of a given sample space S.

The union $A \cup B$ of A and B consists of all points in A or B or both.

The intersection $A \cap B$ of A and B consists of all points that are in both A and B.

If A and B have no points in common, we write

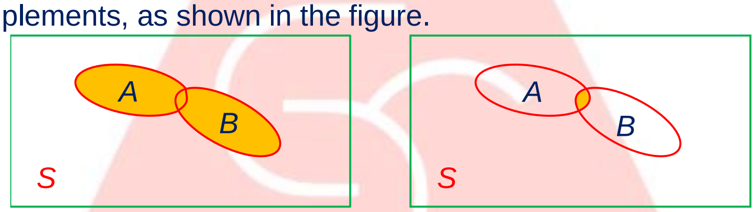
$$A \cap B = \phi$$

where ϕ is the empty set (set with no elements) and we call A and B mutually exclusive (or disjoint) because the occurrence of A excludes that of B (and conversely) if your die turns up an odd number, it cannot turn up an even number in the same trial. Similarly, a coin cannot turn up Head and Tail at the same time.

The complement A^c of A consists of all the points of S not in A. Thus,

$$A \cap A^c = \phi, \quad A \cup A^c = S$$

Working with events can be illustrated and facilitated by Venn diagrams for showing union, intersections, and complements, as shown in the figure.



Union $A \cup B$ Intersection $A \cap B$

Venn diagrams showing two events A and B in a sample space S and their union $A \cup B$ (coloured) and intersection $A \cap B$ (coloured)

Union and intersections of more events are defined similarly. The union

$$\bigcup_{j=1}^{m} A_j = A_1 \cup A_2 \cup \ldots\ldots \cup A_m$$

of events A_1 , ……… , A_m consists of all points that are in at least one A_j. Similarly for the union $A_1 \cup A_2 \cup$ ….. of infinitely many subsets A_1 , A_2 , ……. of an infinite sample space S(that is, S consists of infinitely many points).

The intersection

$$\bigcap_{j=1}^{m} A_j = A_1 \cap A_2 \cap \ldots\ldots \cap A_m$$

of A_1 , ……. , A_m consists of the points of S that are in each of these events. Similarly for the intersection $A_1 \cap A_2 \cap$ ……. of infinitely many subsets of S.

2.6 EQUALLY LIKELY EVENTS

The events are said to be equally likely if none of them is expected to occur in preference to the other one. For example

In throw of a fair coin occurrence of a head or a tail have equal chances. Hence event that a head appears and event that a tail appears are equally likely events.

2.7 MUTUALLY EXCLUSIVE EVENTS

A set of events is said to be mutually exclusive if occurrence of one of them precludes the occurrence of any other. For example:

- In throw of a die, the event of occurrence of an even number and the event of occurrence of an odd number are mutually exclusive.
- In throw of a fair coin, occurrence of a head or a tail are mutually exclusive.

2.8 EXHAUSTIVE EVENTS

A set of events is exhaustive if the performance of the experiment results in occurrence of atleast one of them. For example:

- In throw of a die, the event of occurrence of an even number and the event of occurrence of an odd number are exhaustive.
- In throw of a fair coin, occurrence of a head or a tail is exhaustive.

Exhaustive events cover the whole of the sample space. Their union is equal to *S*.

3 DEFINITION OF PROBABILITY WITH DISCRETE SAMPLE SPACE

If the sample space *S* of an experiment consists of finitely many outcomes (points) that are equally likely, then the probability of occurrence of an event *A* is

$$P(A) = \frac{\text{Number of sample points in A}}{\text{Number of sample points in S}}$$

$$P(A) = \frac{n(A)}{n(S)}$$

In particular $P(S) = 1$ and $0 \le P(A) \le 1$.

4 AXIOMATIC DEFINITION OF PROBABILITY

Given a sample space *S*, with each event A of *S*, there is associated a number *P*(A), called the probability of A, such that the following axioms of probability are satisfied.

- For every A in *S*, $0 \le P(A) \le 1$
- The entire sample space has the probability $P(S) = 1$
- For mutually exclusive events *A* and *B* ($A \cap B = \phi$), $P(A \cup B) = P(A) + P(B)$.

5 BASIC THEORIES OF PROBABILITY

5.1 For an event A and its compliment A^c in sample space *S*

$$\mathbf{P(A^c) = 1 - P(A)} \qquad \text{as } A \cap A^c = \phi \text{ and } A \cup A^c = S$$

and $P(A \cup A^c) = P(A) + P(A^c)$

$\Rightarrow$ $P(S) = P(A) + P(A^c)$

$\Rightarrow$ $1 = P(A) + P(A^c)$

5.2 ADDITION RULE OF PROBABILITY

For events A and *B*

$$\mathbf{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

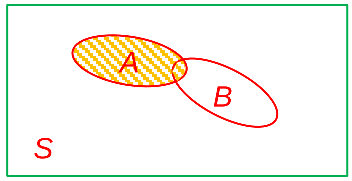
In general for *n* events $A_1, A_2, \ldots\ldots, A_n$ of sample space *S*

$$P(A_1 \cup A_2 \cup \ldots\ldots A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j < n} P(A_i \cap A_j) + \sum_{1 \le i < j < k} P(A_i \cap A_j \cap A_k) \ldots\ldots + (-1)^{n-1} P(A_1 \cap \ldots\ldots \cap A_n)$$

6 CONDITIONAL PROBABILITY

Often it is required to find the probability of an event *B* under the condition that an event A occurs. This probability is called the conditional probability of *B* given A and is denoted by *P*(*B*/A). In this case A serves as a new (reduced) sample space, and the probability is the fraction of that part of set A which corresponds to $A \cap B$. Thus

$$\mathbf{P(B/A) = \frac{P(A \cap B)}{P(A)}} \quad \text{where } P(A) \ne 0)$$



The shaded portion shows the favourable region and lined portion shows the reduced sample space.

Similarly, the conditional probability of A given B is

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

From the above two expressions, we can state the probability of intersection of two events A and B where $P(A) \neq 0$ and $P(B) \neq 0$ as $P(A \cap B) = P(A) \,.\, P\left(\frac{A}{B}\right)$ or $P(B)P\left(\frac{A}{B}\right)$.

(Multiplication theorem)

7 INDEPENDENT EVENTS

If events A and B are such that

$$P(A \cap B) = P(A)\,P(B),$$

they are called independent events. Assuming $P(A) \neq 0$, $P(B) \neq 0$, in this case

$$P(A/B) = P(A)\,, \qquad P(B/A) = P(B)$$

This means that the probability of A does not depend on the occurrence or nonoccurrence of B, and conversely. This justifies the term independent.

Similarly , m events A_1 , ,........ , A_m are called independent if

$$P(A_1 \cap \ldots. \cap A_m) = P(A_1) \ldots\ldots. P(A_m)$$

as well as for every k different events $A_{j_1}, A_{j_2}, \ldots\ldots, A_{j_k}$

$$P(A_{j_1}\cap A_{j_2}\cap \ldots.\cap A_{j_k}) = P(A_{j_1})\,P(A_{j_2})\ldots.P(A_{j_k})$$

where $k = 2, 3, \ldots , m - 1$.

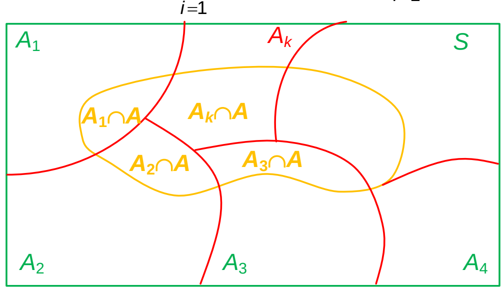
Accordingly, three events A, B, C are independent if

$$P(A \cap B) = P(A)\,P(B),$$
$$P(B \cap C) = P(B)\,P(C),$$
$$P(C \cap A) = P(C)\,P(A)$$
$$P(A \cap B \cap C) = P(A)\,P(B)\,P(C)$$

8 TOTAL PROBABILITY

Consider a sample space S, let A_i, $i = 1$ to n be the set of n mutually exclusive and exhaustive set of sample space S.

Thus $A_i \cap A_j = \phi$ for $1 \leq i < j \leq n$ and $\sum_{i=1}^{n} P(A_i) = 1$ as $\bigcup_{i=1}^{n} A_i = S$



Let A be any event of S. Then total probability of the event A is given by

$$P(A)=\sum_{i=1}^{n}P(A_i)P(A/A_i)$$

where $P(A/A_i)$ gives us the contribution of A_i in the occurrence of A.
This result is obtained as $A = (A_1 \cap A) \cup (A_2 \cap A) \cup \ldots\ldots \cup (A_n \cap A)$

$\Rightarrow$ $P(A) = P(A_1 \cap A) + P(A_2 \cap A) + \ldots\ldots + P(A_n \cap A)$

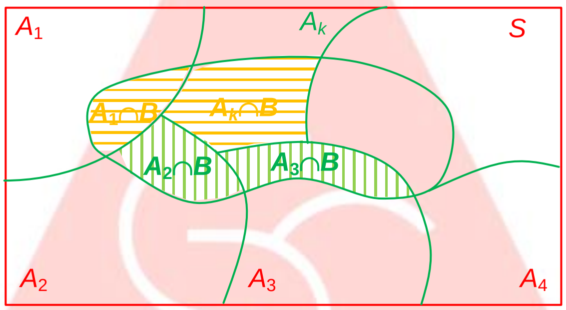
$$= P(A_1)P\left(\frac{A}{A_1}\right) + P(A_2)P\left(\frac{A}{A_2}\right) + \ldots.. + P(A_n)P\left(\frac{A}{A_n}\right) = \sum_{i=1}^{n} P(A_i)P\left(\frac{A}{A_i}\right)$$

9 BAYES THEOREM OR INVERSE PROBABILITY

Bayes theorem gives probability of occurrence of an event when the outcome of experiment is known.

Consider a sample space S, let A_i, $i = 1$ to n be the set of n mutually exclusive and exhaustive set of sample space S.

Thus $A_i \cap A_j = \phi$ for $1 \le i < j \le n$ and $\sum_{i=1} P(A_i) = 1$ as $\bigcup_{i=1}^{n} A_i = S$



Let B be an event of S which has already occurred then conditional probability of occurrence of any one of the event say A_k out of the A_i , i = 1, 2, ….. n events is

$$P\left(\frac{A_k}{B}\right) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)\,P(B/A_k)}{P(B)}$$

Now using the concept of total probability we get Baye's theorem as follows:

$$\boldsymbol{P\left(\frac{A_k}{B}\right) = \frac{P(A_k)\,P(B/A_k)}{\sum_{i=1}^{n} P(A_i)\,P(B/A_i)}}$$

10 RANDOM VARIABLE AND PROBABILITY DISTRIBUTION

A random variable is generally described as a variable whose values are the result of some changing conditions. Consider a simultaneous throw of two coins. The sample space is

$S = \{HH, HT, TH, TT\}$

Let X denote the number of heads in a point of the sample space S. Then

$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0$.

Thus X takes values 0, 1, 2 only and no more. Here we say that X is a random variable or a stochastic variable. We have the following definition.

A random variable is a real valued function *X* defined over the sample space of an experiment i.e. a random variable is a function which associates to each pont of a sample space, a unique real number.

- Random variables are denoted by *X, Y, Z*.
- More than one random variables can be defined on the same sample space.

e.g. Let Y denote the number of heads minus the number of tails for each outcome of the sample space S.

Then $Y(HH) = 2 - 0 = 2$
$Y(HT) = 1 - 1 = 0$
$Y(TH) = 1 - 1 = 0$
$Y(TT) = 0 - 2 = -2$
Thus X and Y are two different random variables defined on the same sample space S.

Probability distribution of random variable

A distribution, in which values of the random variable and their corresponding probabilities are given is called the probability distribution of the random variable.

Let us suppose that a discrete variable X assumes values $x_1, x_2, \ldots\ldots x_n$ with probability $p_1, p_2 \ldots\ldots p_n$ respectively, where $p_1 + p_2 + \ldots\ldots + p_n = 1$ and $0 \le p_i \le 1$ for all $i = 1, 2, \ldots .n$.

Then the following table describes the probability distribution.

X	x_1	x_2	x_3	x_4	………	x_n
$P(X)$	p_1	p_2	p_3	p_4	………	p_n

Example:

Let X be the random variable denoting the number of tails in a simultaneous throw of two coins. Then clearly X can take the values 0, 1, 2.

$X(TT) = 2,\ X(HT) = 1,\ X(TH) = 1$ and $X(HH) = 0$

Let $P(X) =$ Prob. of the variable X.

Then $P(X = 0) = P\,(\text{no tail}) = \frac{1}{4}$

$P(X = 1) = P\ (\text{one tail}) = \frac{2}{4} = \frac{1}{2}$

$P(X = 2) = P\,(\text{two tails}) = \frac{1}{4}$

Note:

$$P(X = 0) + P(X = 1) + P(X = 2) = 1 \quad \left[\because \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1\right]$$

We can write the above result in the following form:

X	0	1	2
$P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Clearly, each of the probability is a non-negative fraction (never greater than 1) and their sum is 1. The above form is the probability distribution of the random variable X.

11 BINOMIAL DISTRIBUTION

In n independent trials of a random experiment, let X be the number of times an event A occurs. In each trial, event A has same probability as $P(A) = p$ referred to as success. Then in a trial non-occurrence of A is referred as failure and given by $q = 1 - p$.

Here X can assume values from o to n. Now $X = r$ means A occurs in r trials and $(n - r)$ it does not occur this may look as

$$\underbrace{A\ A\ A \ldots\ldots A}_{r \text{ times}} \quad \underbrace{\bar{A}\ \bar{A} \ldots\ldots \bar{A}}_{n-r \text{ times}}$$

here $\bar{A}$ means complement of A. Using the assumption that trials are independent, that is, they do not influence each other, hence has the probability

$$\underbrace{p\ p\ ....p}_{r\text{ times}}\ \underbrace{q\ q.........q}_{n-r\text{ times}} = p^r\ q^{n-r}$$

and it can be arranged in $\dfrac{n!}{r!(n-r)!} = {}^nC_r$ ways.

Hence the probability of getting r successes or occurrence of A in r trials out of n independent trials is

$$\mathbf{P(X = r) = {}^nC_r\, p^r\, q^{n-r}}$$

which is denoted as binomial distribution of random variable X.

- The probability of getting atleast k successes in

$$P(X \geq k) = \sum_{r=k}^{n} {}^nC_r p^r q^{n-r}$$

- The probability of getting almost K successes is

$$P(X \leq k) = \sum_{r=0}^{k} {}^nC_r p^r q^{n-r}$$

- $$\sum_{r=0}^{n} {}^nC_r p^r q^{n-r} = (p+q)^n = 1$$

Recurrence Formula for Binomial Distribution

We know that $P(r) = {}^nC_r q^{n-r} p^r$ [**note:** $P(r)$ means $P(X = r)$]

$$P(r+1) = {}^nC_{r+1} q^{n-r-1} p^{r+1}$$

$$\frac{P(r+1)}{P(r)} = \frac{{}^nC_{r+1}}{{}^nC_r} \cdot \frac{q^{n-r-1}}{q^{n-r}} \cdot \frac{p^{r+1}}{p^r}$$

$$= \frac{n-r}{r+1} \cdot \frac{p}{q}$$

Hence $P(r+1) = \dfrac{n-r}{r+1} \cdot \dfrac{p}{q} P(r)$,

which is the required recurrence formula for binomial distribution.

- If $P(0)$ is known, then we find $P(1)$, $P(2)$, …… with the help of recurrence formula.
- If $p = \frac{1}{2}$, then $q = \frac{1}{2}$. Then the binomial distribution is called symmetrical binomial distribution.

Mean and Variance of binomial distribution

If $x_1, x_2, ...x_n$ are the values of a random variable X and $p_1, p_2, ...p_n$ are the corresponding probabilities, then mean (μ) and variance (σ^2) of the probability distribution are given by

$$\mu = \Sigma p_i x_i$$

$$= 0.p(0) + 1.p(1) + .2p(2) + ... + n.p(n)$$

$$= 1.\ {}^nC_1 pq^{n-1} + 2.\ {}^nC_2 p^2 q^{n-2} + ... + n\ {}^nC_n p^n$$

$$= npq^{n-1} + 2\frac{n(n-1)}{2}.p^2 q^{n-2} + 3.\frac{n(n-1)(n-2)}{1.2.3}.p^3 q^{n-3} + ... + n.p^n$$

$$= np[q^{n-1} + (n-1)pq^{n-2} + \frac{(n-1)(n-2)}{1.2} p^2 q^{n-3} + ... + p^{n-1}]$$

$$= np[\ {}^{n-1}C_0 q^{n-1} + C_1 pq^{n-2} + {}^{n-1}C_2 p^2 q^{n-3} + ... + {}^{n-1}C_{n-1} p^{n-1}]$$

$= np(q+p)^{n-1} = np(1)^{n-1}$ $[\because q+p=1]$

$= np$

Hence the mean of the distribution i.e. $\mu = np$

Again variance σ^2 is given by $\sigma^2 = \Sigma p_i x_i^2 - \mu^2$

$= 0^2 p(0) + 1^2.p(1) + 2^2.p(2) + ... + n^2.p(n) - \mu^2$

$= 1.\ {}^nC_1 pq^{n-1} + 2^2.\ {}^nC_2 p^2 q^{n-2} + ... + n^2.\ {}^nC_n p^n - \mu^2$

$= [npq^{n-1} + 4\dfrac{n(n-1)}{1.2}p^2q^{n-2} + 9.\dfrac{(n-1)(n-2)}{1.2.3}p^3q^{n-3} + ... + n^2.p^n] - \mu^2$

$= np[q^{n-1} + 2.(n-1)pq^{n-2} + 3.\dfrac{(n-1)(n-2)p^2}{1.2}q^{n-3} + ... + np^{n-1}] - \mu^2$

$= np[\ {}^{n-1}C_0 q^{n-1} + 2^{(n-1)}C_1 p^1 q^{n-2} + 3.\ {}^{(n-1)}C_2 p^2 q^{n-3} + ... + np^{n-1}] - \mu^2$

$= np[(\ {}^{n-1}C_0 q^{n-1} + {}^{n-1}C_1 pq^{n-2} + {}^{n-1}C_2 p^2 q^{n-3} + ... + {}^{n-1}C_{n-1} p^{n-1})]$

$+ (\ {}^{n-1}C_1 pq^{n-2} + 2.\ {}^{n-1}C_2 p^2 q^{n-3} + ... + (n-1.\ {}^{n-1}C_{n-1} p^{n-1})] - \mu^2$

$= np[(q+p)^{n-1} + (n-1)p(\ {}^{n-2}C_0.q^{n-2}) + {}^{n-2}C_1 pq^{n-3} + ... + {}^{n-2}C_{n-2} p^{n-2})] - \mu^2$

$$\left[\because r.\frac{{}^{n-1}C_r}{n-r} = {}^{n-1}C_{r-1} \text{ for } r = 1, 2, 3....n-1\right]$$

$= np[(1)^{n-1} + (n-1)p.(q+p)^{n-2}] - n^2p^2$ $[\because \mu = np]$

$= np[1 + (n-1)p.(1)^{n-2}] - n^2p^2$

$= np[1 + (n-1)p] - n^2p^2 = np[1 + np - p] - n^2p^2$

$= np[np + q] - n^2p^2$ $[\because 1 - p = q]$

$= n^2p^2 + npq - n^2p^2 = npq$

Hence Variance of Binomial Distribution is given by $\sigma^2 = npq$

- Standard Deviation (S.D.) for the Binomial Distribution $= \sigma = \sqrt{npq}$
- Variance of Binomial Distribution is less than its mean

$\because$ Variance $= npq \le np$

Mean = np

Hence *Variance* $\le$ *Mean of the Binomial Distribution*.

i.e. mean of the binomial distribution is always greater than the variance.

12 POISSON DISTRIBUTION

It is limiting case of binomial distribution. If the number of events n is very large $(n \to \infty)$ and probability of success in each experiment is p ('p' being very small) and $np = \lambda$ (say) is finite, then

$P(X = r)$ or $P(r) = \dfrac{e^{-\lambda}\lambda^r}{r\,!}$, where $r = 0, 1, 2, ...$ and $\lambda = np$. Here λ is known or parameter of distribution.

$P(r+1) = \dfrac{\lambda}{r+1}P(r)$ is known as recurrence formula.

Note:

- $\sum_{r=0}^{\infty} P(r) = 1$
- If λ_1 and λ_2 are parameter of variables X and Y, then parameter of X+Y will be $(\lambda_1 + \lambda_2)$.
- In poisson distribution, mean = variance = λ .

13 PROBABILITY OF EVENTS IN EXPERIMENTS WITH COUNTABLE INFINITE SAMPLE SPACE

The sample space of some random experiments have infinite sample points, and hence the events also have infinite sample points in favour of their occurrence.

In some cases the probability is calculated by relating the sample space or the sets representing the events, with lengths or areas of geometrical figures etc. Following are the illustrations of such cases.