Note for module-5. (Applied Probability).

Definitions:-

Random Experiment:- An experiment E is called random if (i) all possible outcomes of E are known in advance, (ii) It is impossible to predict which outcome will occur at a particular performance of E & (iii) E can be repeated, at least conceptually, under identical conditions for infinite number of times.

Example: Throwing a dice, Tossing coin etc.

Sample space/ Event space:- The set of all possible outcomes connected to a random experiment.

Ex:- $S = \{1,2,3,4,5,6\}$ for dice throwing.

$S = \{HH, HT, TH, TT\}$ for tossing a coin twice successively

Event:- Any member of a sigma algebra (σ- Algebra) over the event space S is called an event.

Clarification:-

σ-Algebra over a set S:- It is a collection (Δ) of subsets of S (or in other words a subset of powerset of S (set of all subsets of S)) is called a sigma algebra if.

(i) $S \in \Delta$

(ii) If $A \in \Delta$, then $A^c \in \Delta$

(iii) If $A_1, A_2, \ldots, A_n, \ldots \in \Delta$ (countable collection)

Then $\bigcup_{i=1}^{\infty} A_i \in \Delta$.

Ex:- (a) 2^S (power set of any set S) is the ~~trivial~~ improper/trivial sigma algebra over any set S.

(b) $\Delta = \{\emptyset, S\}$ is also a trivial sigma algebra over any set S.

(c) For $S = \{1,2,3,4,5,6\}$
$\Delta = \{\emptyset, S, \{1,3,5\}, \{2,4,6\}\}$ is a proper / non trivial sigma algebra over S.
(Check that all properties are satisfied).

[Note:- So for defining event for a random experiment we first need a sigma algebra over the sample space (if nothing mentioned we will consider the sigma algebra mentioned in example (a)) & thus any subset E of sample space S will become a event if nothing mentioned.]

Simple event:- An event E connected to a random experiment A with associate sample space S & a σ- algebra Δ is called simple if $|E| = 1$ i.e only one element in E.

Ex:- For coin tossing (twice simultaneously) $E = \{HH\}$ is a simple event. (We have taken whole power set of $S = \{HH, HT, TH, TT\}$ to be our sigma algebra here).

composite/ ~~Compound Event~~:- Event E which is not simple is compound / composite. i.e. $|E| > 1$.

Ex:- For dice throwing the set of even outcomes i.e $A = \{2,4,6\}$ is a composite event consisting 3 simple events $\{2\}$, $\{4\}$, $\{6\}$.

2.6. Mutually Exclusive Events.

Two events connected to a given random experiment E are said to be mutually exclusive if A, B can never happen simultaneously in any performance of E, *i.e.*, if $AB = O$. In connection with the random experiment of throwing a die, the events 'multiple of 3' and 'a prime number' are not mutually exclusive, since the number '3' is a multiple of 3 as well as a prime number, whereas the events 'even number' and 'odd number' are mutually exclusive events of the same random experiment.

2.7. Exhaustive Set of Events.

A collection of events is said to be exhaustive if in every performance of the corresponding random experiment at least one event (not necessarily the same for every performance) belonging to the collection happens. In set theoretic notations the collection of events $\{A_\alpha : \alpha \in I\}$ is exhaustive if and only if

$$\sum_{\alpha \in I} A_\alpha = S,$$

where I is an index set and S is the corresponding event space. In connection with the random experiment of throwing a die the collection of events $\{A_1, A_2, A_3\}$ is exhaustive where

$$A_1 = \{1, 3, 5\}, \quad A_2 = \{2\}, \quad A_3 = \{4, 6\}.$$

2.8. Statistical Regularity.

Let a random experiment E be repeated N times under identical conditions, in which we note that an event A of E occurs $N(A)$

times. Then the ratio $\frac{N(A)}{N}$ is called the frequency ratio of A and is denoted by $f(A)$. Now if the random experiment E is repeated a very large number of times, it is seen that the frequency ratio $f(A)$ gradually stabilises to a more or less constant, *i.e.*, $f(A)=\frac{N(A)}{N}$ gradually tends to a constant number as N becomes larger and larger. This tendency of stability of frequency ratio is called statistical regularity and this fact was confirmed by many experimental results.

2.9. Classical Definition of Probability.

Towards the beginning of the 19th century, Laplace gave a formal definition of probability which goes by the name of the classical definition. The theory of probability developed on the basis of the classical definition is known as the classical theory of probability. In the classical theory we have the following definition of probability :

Let the event space S of a given random experiment E be finite. If all the simple events connected to E be 'equally likely' then the probability of an event $A(A \subseteq S)$ is defined as

$$P(A)=\frac{m}{n},$$

where n is the total number of simple events connected to E, *i.e.*, n is the number of distinct elements of S and m of these simple events are favourable to A, *i.e.*, A contains m distinct elements.

At this stage it is not possible to give a precise definition of the phrase 'equally likely' used in the above definition. In the next section we shall critically examine the meaning of the phrase. At present, we shall say that all the simple events are equally likely if it is understood intuitively that no one of them is expected to occur in preference to others in any trial of the given random experiment and only then the definition can be applied.

Deductions :—

(*a*) $0 \leq P(A) \leq 1$

(*b*) $P(S)=1$

(c) $P(O)=0$

(d) $P(\bar{A})=1-P(A)$.

Proof: (a) We have $P(A)=\frac{m}{n}$, where m, n have the meanings given before.

$$\text{Here} \quad 0 \le m \le n$$
$$\text{or,} \quad 0 \le \frac{m}{n} \le 1.$$
$$\text{So,} \quad 0 \le P(A) \le 1.$$

(b) $P(S)=\frac{n}{n}=1$.

(c) $P(O)=\frac{0}{n}=0$.

(d) $P(\bar{A})=\frac{n-m}{n}=1-\frac{m}{n}=1-P(A)$.

THEOREM 2.9.1. Theorem of Total Probability.

If $A_1, A_2, \dots, A_k$ are pairwise mutually exclusive events, then
$P(A_1+A_2+\dots+A_k)=P(A_1)+P(A_2)+\dots+P(A_k)$.

Proof: Let n be the total number of simple events of the corresponding random experiment E of which m_i are favourable to A_i, $i=1, 2, \dots, k$. Since the events $A_1, A_2, \dots, A_k$ are mutually exclusive, the total number of simple events favourable to the event $A_1+A_2+\dots+A_k$ is $m_1+m_2+\dots+m_k$,

$$0 \le m_i \le n, \; i=1, 2, \dots, k.$$

Then by the classical definition,

$$P(A_1+A_2+\dots+A_k)=\frac{m_1+m_2+\dots+m_k}{n}$$
$$=\frac{m_1}{n}+\frac{m_2}{n}+\dots+\frac{m_k}{n}$$
$$=P(A_1)+P(A_2)+\dots+P(A_k).$$

Hence the theorem.

2.10. Criticisms of the Classical Definition.

If we examine the classical definition a little more closely we find that there is a logical drawback in the definition. We note that the definition can be used only if it is possible to ascertain that all the simple events are equally likely. In many problems, considerations of symmetry and similarity enable us to decide whether, in the problem before us, simple events are equally likely. For example, if a die be symmetric, then the simple events connected to the random experiment of throwing the die may be considered to be equally likely. But it is very difficult to explain the nature of 'symmetry' and 'similarity' as stated above. It was found after many serious investigations that the phrase 'equally likely' cannot be explained without the prior idea of probability.

Moreover, the definition is restricted to event spaces which are finite and where all the simple events are equally likely. The definition cannot be applied where the simple events are not equally likely or where the event space is infinite. With the help of this definition it will thus be impossible to treat the case of a loaded die since here intuitively we can expect that a face can turn up in preference to others and consequently simple events are not necessarily equally likely and the case of predicting the number of telephone calls in a given interval (in a given trunk line) in which there are infinite number of simple events.

In order to avoid the limitations of the classical approach and to make the definition more widely applicable, we now take recourse in the next section to another definition, called the frequency definition of probability.

It may be noted here that the classical definition is based on advance subjective concept of probability so that $P(A)=\frac{m}{n}$ should rather be called a method of calculation or probability for events of a finite event space of equally likely simple events, instead of taking it as a definition of probability.

2.11. Frequency Definition of Probability.

Let A be an event of a given random experiment E. Let the event A occur $N(A)$ times when the random experiment E is repeated N times under identical conditions. Then on the basis of statistical regularity we can assume that $\underset{N\to\infty}{Lt} \frac{N(A)}{N}$ exists finitely and the value of this limit is called the probability of the event A, denoted by $P(A)$,

$$\text{i.e., } P(A) = \underset{N\to\infty}{Lt} \frac{N(A)}{N} = \underset{N\to\infty}{Lt} f(A),$$

where $f(A) = \dfrac{N(A)}{N}$ is the frequency ratio of the event A in N repetitions of the corresponding random experiment under identical conditions.

Deductions :

(a) $0 \leq P(A) \leq 1$, for any event A

(b) $P(S) = 1$

(c) $P(O) = 0$

(d) $P(\bar{A}) = 1 - P(A)$.

Proof : (a) We have $0 \leq N(A) \leq N$, where N and $N(A)$ have the meanings given above.

$$\therefore \; 0 \leq \frac{N(A)}{N} \leq 1$$

$$\text{or, } 0 \leq \underset{N\to\infty}{Lt} \frac{N(A)}{N} \leq 1.$$

Hence, $0 \leq P(A) \leq 1$.

(b) $$P(S) = \underset{N\to\infty}{Lt} \frac{N(S)}{N} = \underset{N\to\infty}{Lt} \frac{N}{N} = 1.$$

(c) $$P(O) = \underset{N\to\infty}{Lt} \frac{N(O)}{N} = \underset{N\to\infty}{Lt} \frac{0}{N} = 0.$$

(d) $$P(\bar{A}) = \underset{N\to\infty}{Lt} \frac{N(\bar{A})}{N}$$

$$= \underset{N\to\infty}{Lt} \frac{N - N(A)}{N}$$

$$= \underset{N\to\infty}{Lt} \left[1 - \frac{N(A)}{N}\right]$$

$$= 1 - \underset{N\to\infty}{Lt} \frac{N(A)}{N}$$

Definition of conditional probability.

Let E be a given random experiment and A, B be two events of E where $P(B) \neq 0$. The conditional probability of the event A on the hypothesis that the event B has happened, denoted by $P(A \mid B)$, is defined by

$$P(A \mid B) = \underset{N \to \infty}{Lt} \frac{N(AB)}{N(B)},$$

assuming that the limit exists, N, $N(AB)$, $N(B)$ have the usual meanings given before.

THEOREM 2.12.1. Theorem of Compound Probability.

If A, B are two events of a given random experiment, then

$$P(AB) = P(A/B)\, P(B), \text{ if } P(B) \neq 0$$

$$\text{or, } P(AB) = P(B/A)\, P(A), \text{ if } P(A) \neq 0.$$

Proof: Let E be the given random experiment and let E be repeated under identical conditions N times. If $N(AB)$, $N(B)$, $N(A)$ be respectively the number of occurrences of the events AB, B, A, then

$$P(A/B) = \underset{N \to \infty}{Lt} \frac{N(AB)}{N(B)}$$ [Here we note that $P(B) > 0 \Rightarrow N(B) > 0$ for suitable large N.]

$$= \underset{N \to \infty}{Lt} \frac{\dfrac{N(AB)}{N}}{\dfrac{N(B)}{N}}$$

$$\text{or,} \quad P(A/B) = \frac{\underset{N\to\infty}{Lt} \frac{N(AB)}{N}}{\underset{N\to\infty}{Lt} \frac{N(B)}{N}}$$

$$= \frac{P(AB)}{P(B)}. \quad \left(\because \underset{N\to\infty}{Lt} \frac{N(B)}{N} = P(B) \neq 0 \right).$$

Hence we get,

$$P(AB) = P(A \mid B)\, P(B) \text{ if } P(B) \neq 0.$$

It can be proved similarly that

$$P(AB) = P(B \mid A)\, P(A) \text{ if } P(A) \neq 0.$$

Remark : We know that the event S occurs in any trial of a given random experiment, so from the meaning of conditional probability given above we find that unconditional probability $P(A)$ is a particular conditional probability, since the statement 'A occurs' can also be expressed as 'A occurs on the hypothesis that S has happened'. So $P(A \mid S)$ and $P(A)$ should be equal and by the theorem of compound probability we find that

$$P(A \mid S) = \frac{P(AS)}{P(S)} = \frac{P(A)}{1} = P(A).$$

2·13. Criticisms of the Frequency Definition.

In this definition we note that the frequency ratio $\frac{N(A)}{N}$ is obtained from observation whereas $\underset{N\to\infty}{Lt} \frac{N(A)}{N}$ is a rigorous analytical concept. This combination of empirical and analytical concepts leads to mathematical difficulties. Although there is not much objection against the logical content of the theory of probability based on the frequency definition but due to the aforesaid weakness in the definition it will be unwise to build the theory of probability on the basis of this definition.

Conclusion : In sections 2.10 and 2.13 we have seen that classical and frequency definitions are both inadequate for developing the mathematical theory of probability. Now the theory of probability is conceived as a mathematical theory of phenomena showing statistical regularity. So in order that mathematical theory of probability may be applied to different types of phenomena showing statistical regularity, the definition of probability should be independent of the intended application. From all these considerations we feel the necessity of an axiomatic treatment of the theory of probability, *i.e.*, the theory of probability, as a branch of mathematics, should be developed from axioms in exactly the same way as Geometry and Algebra. Axioms are propositions which are regarded as true and not proved within the framework of the given theory. All other propositions of the theory have to be proved from the accepted axioms in a purely logical manner. Axiomatic theory starts from one or more sets of abstract objects, where some relations between the objects are expressed by the axioms. The points, lines, planes considered in Pure Geometry as abstract objects, are not things that we know from immediate experience. Pure Geometry deals with such abstract objects entirely defined by their properties, as expressed by the five sets of axioms, namely, 'axioms of incidence', 'axioms of order', 'axioms of motion', 'axiom of parallelism' and 'axiom of continuity' (Hilbert).

Now any mathematical theory developed logically from a set of axioms can have many concrete interpretations besides those from which the axioms are developed. Similar is the situation in the axiomatic theory of probability. But we shall interpret the theory in such a way that 'events' will be events of the real world and the probability will be so interpreted that it can be applied to phenomena showing statistical regularity.

Formulation of the axioms are the results of a prolonged accumulation of facts and a logical analysis of the results obtained and in this way the axioms of Geometry, studied in elementary mathematics, were formulated. The axioms taken for defining probability will certainly be motivated by the results obtained from the classical and frequency definition of probability. On the basis of the axioms it will be possible to construct a logically consistent theory of probability. We shall begin the next chapter with the axioms proposed by the Russian mathematician A. N. Kolmogorov.

Let E be a given random experiment and S be the corresponding event space. Also let Δ be the class of subsets of S forming the class of events of E. A mapping $P : \Delta \to R$ is called a probability function defined on Δ and the unique real number $P(A)$ determined by P is called the probability of the event A where $A \in \Delta$ if the following axioms, known as axioms of probability, are satisfied :

Axiom (*a*). $P(A) \geqslant 0$ for every event $A \in \Delta$.

Axiom (*b*). $P(S) = 1$.

Axiom (*c*). If $A_1, A_2, \ldots\ldots, A_n, \ldots$ be countably infinite number of pairwise mutually exclusive events, *i.e.*, if $A_i A_j = O$ whenever $i \neq j$ and $A_i, A_j \in \Delta$,

$$\text{then } P(A_1 + A_2 + A_3 + \cdots + A_n + \cdots) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots \quad (3.1.1)$$

The entire mathematical theory of probability will be built by three objects, namely (*i*) the event space (*ii*) the class of events Δ (*iii*) the probability function $P : \Delta \to R$. The ordered 3 – *tuple* (S, Δ, P) is called a probability space.

• Properties of probability function (axiomatic definition)

① $P(\emptyset) = 0$

② If $A_1, A_2, \ldots, A_n$ be finite number of pairwise mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

or equivalently denoted by

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

(sum b/w sets represent their union & multiplication represent their intersection).

③ $P(\bar{A}) = 1 - P(A)$

↓ complement of A

④ Classical definition can be derived from axiomatic definition i.e. axiomatic definition doesn't violate intutive classical definition.

⑤ $0 \leq P(A) \leq 1$ for any event A.

⑥ If $A \subseteq B \Rightarrow P(A) \leq P(B)$.

⑦ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Addition rule)

or $P(A+B) = P(A) + P(B) - P(AB)$

General addition rule is given by

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i_1, i_2 = 1 \\ (i_1 < i_2)}}^{n} P(A_{i_1} A_{i_2}) + \cdots$$

$$+ \sum_{\substack{i_1, i_2, i_3 = 1 \\ i_1 < i_2 < i_3}}^{n} P(A_{i_1} A_{i_2} A_{i_3}) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

THEOREM 3.10.2. **Bayes' Theorem.**

Let $A_1, A_2, \ldots\ldots, A_n$ be n pairwise mutually exclusive events connected to a random experiment E where at least one of $A_1, A_2, \ldots, A_n$ is sure to happen (i.e., $A_1, A_2, \ldots\ldots, A_n$ form an exhaustive set of n events). Let X be an arbitrary event connected to E, where $P(X) \neq 0$. Also let the probabilities $P(X \mid A_1), P(X \mid A_2), \ldots\ldots, P(X \mid A_n)$ be all known.

Then $$P(A_i \mid X) = \frac{P(A_i)P(X \mid A_i)}{\sum_{r=1}^{n} P(A_r)P(X \mid A_r)}, \quad i = 1, 2, \ldots\ldots, n. \tag{3.10.2}$$

3.11. Independence of Events.

Let A and B be two events connected to a given random experiment. If $P(B) \neq 0$ then $P(A \mid B)$ can be defined and in this case if $P(A \mid B) = P(A)$, then we can say that the probability of A does not depend on the happening of B, *i.e.*, there is one kind of independence between A and B. Also if $P(A) \neq 0$, then $P(B \mid A)$ can be defined and in this case if $P(B \mid A) = P(B)$, we can say that the probability of B does not depend on the happening of A, *i.e.*, there is one kind of independence between A and B. We observe that

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

both lead to $P(AB) = P(A)\, P(B)$.

So formally we can define independence of two events as follows :

Two events A, B are said to be stochastically independent or statistically independent or simply independent if and only if

$$P(AB) = P(A)P(B). \qquad (3.11.1)$$

If $P(AB) \neq P(A)P(B)$, then A, B are said to be dependent.

3.12. Mutual and Pairwise Independence of more than two Events.

Three events A, B, C are said to be pairwise independent if

$$\left.\begin{aligned} P(AB) &= P(A)P(B) \\ P(BC) &= P(B)P(C) \\ P(CA) &= P(C)P(A) \end{aligned}\right\} \tag{3.12.1}$$

and A, B, C are said to be mutually independent if

$$\left.\begin{aligned} P(AB) &= P(A)P(B) \\ P(BC) &= P(B)P(C) \\ P(CA) &= P(C)P(A) \\ P(ABC) &= P(A)P(B)P(C). \end{aligned}\right\} \tag{3.12.2}$$

In general n events $A_1, A_2, \ldots\ldots, A_n$ $(n > 2)$ are said to be mutually independent if

$P(A_iA_j) = P(A_i)P(A_j)$, where $i < j$; i, j any combination of $1, 2, \ldots\ldots, n$ taken two at a time.

$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k)$, where $i < j < k$; i, j, k any combination of $1, 2, \ldots\ldots, n$ taken 3 at a time.

... ...

... ...

$$P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n). \tag{3.12.3}$$

Note 1. From (3.12.3) we see that in defining mutual independence of n events $(n > 2)$,

$$^nC_2 + {}^nC_3 + \cdots + {}^nC_n = 2^n - n - 1$$

relations are required.

Note 2. From the definition of mutual independence, we see that mutual independence implies pairwise independence, but the converse is not true, as shown by the following example.

Let the equally likely outcomes of an experiment be one of the four points in the three-dimensional space with rectangular co-ordinates (1, 0, 0), (0, 1, 0), (0, 0, 1) and (1, 1, 1). Let A, B, C denote the events 'x-co-ordinate 1', 'y-co-ordinate 1' and 'z-co-ordinate 1' respectively.

Then by using classical definition,

$$P(A)=\tfrac{2}{4}=\tfrac{1}{2},\quad P(B)=\tfrac{1}{2},\quad P(C)=\tfrac{1}{2}.$$
$$P(AB)=\tfrac{1}{4}=P(A)P(B)$$
$$P(BC)=\tfrac{1}{4}=P(B)P(C)$$
$$P(CA)=\tfrac{1}{4}=P(C)P(A).$$

Hence, A, B, C are pairwise independent.

But $P(ABC)=\tfrac{1}{4}$.

$\therefore\ P(ABC)\neq P(A)P(B)P(C)$,

which implies that A, B, C are not mutually independent.

Hence, pairwise independence does not always imply mutual independence.

Note 3. It is to be noted that the concept of mutually exclusive events and independent events are not equivalent. We bring out the difference between the two ideas.

If two events A and B are mutually exclusive then $AB=O$ and so the occurrence of one of the two events, in this case, is hindered by anticipating the occurrence of the other.

On the other hand, if the occurrence of one event has no effect on the probability of the occurrence of the other event, the two events are said to be independent and in this case $P(AB)=P(A)P(B)$.

Two events can be mutually exclusive and not independent. For example, consider the random experiment of tossing of two coins. Let A and B be the events 'both the coins show head' and 'both the coins show tail' respectively. Then A and B are clearly mutually exclusive, since if A happens B cannot happen and as such $AB=O$. But $P(A)=\tfrac{1}{4}$, $P(B)=\tfrac{1}{4}$. $\therefore\ P(AB)=0\neq P(A)P(B)$, *i.e.*, A and B are not independent.

Again, two events can be independent and not mutually exclusive. For example, consider the random experiment of throwing 2 dice together. Let A and B be the events '6 appears in the first die' and '6 appears in the second die' respectively.

Then $P(AB)=\tfrac{1}{36}=P(A)P(B)=\tfrac{1}{6}\times\tfrac{1}{6}$ and so A and B are independent. Also $AB=\{(6,6)\}\neq O$ which implies that A and B are not mutually exclusive.

Finally, two events A and B can be both mutually exclusive and independent when

$$P(AB)-P(A)P(B)=0,$$

which holds if at least one of the two events A and B has zero probability. In fact, two events having both non-zero probabilities cannot be simultaneously mutually exclusive and independent.

3.13. General Multiplication Rule.

THEOREM 3.13.1. *If $A_1, A_2, \ldots\ldots, A_n$ $(n \geqslant 2)$ be n events connected to a random experiment E, then*

$$P(A_1A_2\ldots A_n)=P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\ldots P(A_n \mid A_1A_2\ldots A_{n-1}) \quad (3.13.1)$$

provided the conditional probabilities are defined.

Proof: For two events A_1, A_2 we have, by the definition of conditional probability,

$$P(A_2 \mid A_1)=\frac{P(A_1A_2)}{P(A_1)}.$$

$$\therefore \quad P(A_1A_2)=P(A_1)P(A_2 \mid A_1).$$

Hence the proposition (3.13.1) is true for $n=2$.

Let the proposition be true for $n=m$, where m is a positive integer $\geqslant 2$. Then we have for any m events $A_1, A_2, \ldots A_m$,

$$P(A_1A_2A_3\ldots A_m)=P(A_1)P(A_2 \mid A_1)\ldots P(A_m \mid A_1A_2\ldots A_{m-1}). \quad (3.13.2).$$

Now we consider the $(m+1)$ events $A_1, A_2, \ldots, A_m, A_{m+1}$.

Then $P(A_1A_2\ldots\ldots A_mA_{m+1})$

$=P[(A_1A_2\ldots\ldots A_m)A_{m+1}]$

$=P(A_1A_2\ldots\ldots A_m)P(A_{m+1} \mid A_1A_2\ldots\ldots A_m)$,

since the proposition is true for $n=2$

$=P(A_1)P(A_2 \mid A_1)\ldots\ldots P(A_m \mid A_1A_2\ldots\ldots A_{m-1})$
$P(A_{m+1} \mid A_1A_2\ldots\ldots A_m)$,

by (3.13.2)

This shows that the proposition (3.13.1) is true for $n=m+1$ if it is true for $n=m$. But the proposition is true for $n=2$. Hence by the principle of mathematical induction, the proposition is true for any positive integer $n \geqslant 2$.

Particular Case.

For three events A_1, A_2, A_3 connected to a random experiment E,

$$\begin{aligned} P(A_1A_2A_3) &= P(BA_3), \quad \text{where } B=A_1A_2 \\ &= P(B)P(A_3|B) \\ &= P(A_1A_2)P(A_3 \mid A_1A_2) \\ &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2). \end{aligned}$$

• **Random Variable:-** Let Δ be a sigma algebra over S, sample space, connected to the random experiment E. Then a mapping $X: S \to \mathbb{R}$ is called a random variable if

$$X^{-1}((-\infty, x]) \in \Delta \quad \forall\, x \in \mathbb{R}$$

i.e. $\{\omega : -\infty < X(\omega) \leq x, \omega \in S\}$ is an event connected to E $\forall\, x \in \mathbb{R}$.

The range of X is known as spectrum of the random variable X.

[Note: After recalling definition of sigma algebra & exploiting its inner meaning by yourself, and observing the definition of random variable carefully, try to show that the sets $\{\omega : X(\omega) = x, \omega \in S\}$, $\{\omega : x < X(\omega) < \infty, \omega \in S\}$, $\{\omega : -\infty < X(\omega) < x, \omega \in S\}$, $\{\omega : a < X(\omega) < b, \omega \in S\}$, $\{\omega : X(\omega) \geq c, \omega \in S\}$, etc all are events whenever X is a random variable.

In future, we will write the events

$\{\omega : -\infty < X(\omega) \leqslant x, \omega \in S\}$, $\{\omega : X(\omega)=x, \omega \in S\}$,
$\{\omega : a < X(\omega) \leqslant b, \omega \in S\}$ and so on, in short as
$(-\infty < X \leqslant x)$, $(X=x)$, $(a < X \leqslant b)$ respectively and so on.

Ex. A coin is tossed twice. Here,

$$S=\{\omega_1=(H, H), \omega_2=(H, T), \omega_3=(T, H), \omega_4=(T, T)\}.$$

A mapping $X : S \to R$ is defined as follows :

$X(\omega_i)=k$, where k is the number of heads, $i=1, 2, 3, 4.$

Then $X(\omega_1)=2$, $X(\omega_2)=X(\omega_3)=1$, $X(\omega_4)=0$. Here X is a random variable defined in the domain S and the spectrum (range) of X is $\{0, 1, 2\}$. Here, according to our notation $(X=0)$ represents the event $\{(T, T)\}$, $(0 \leqslant X \leqslant 2)$ is a certain event and $(1 < X < 2)$ represents the impossible event O.

The above random variable $X : S \to R$ is also described in the following manner. The random variable X, in this case, defined on S denotes the total number of heads in two tosses of the coin. Later, we shall often use this convention of description of a random variable.

5·2. Distribution Function.

Let $P : \Delta \to R$ be a probability function, where Δ is the class of subsets (of S) forming the class of events. We remember that, the ordered 3 tuple (S, Δ, P) is called a *probability space.*

Let X be a random variable defined on the event space S connected to a random experiment E. The *distribution function* of the random variable X with respect to the probability space (S, Δ, P) is a real valued function $F(x)$ of a real variable x, defined in $(-\infty, \infty)$, where

$$F(x)=P(-\infty < X \leqslant x), \text{ for all } x \in (-\infty, \infty). \qquad (5.2.1)$$

It is evident that the range of the distribution function is a subset of $[0, 1]$.

• Properties of Distribution Function :

① $0 \le F(x) \le 1 \quad \forall x \in (-\infty, \infty)$

since $0 \le P(-\infty < X \le x) \le 1 \quad \forall x \in (-\infty, \infty)$

$\Rightarrow 0 \le F(x) \le 1 \quad \forall x \in (-\infty, \infty)$

② $P(a < X \le b) = F(b) - F(a)$

As the events $(-\infty < X \le a)$ & $(a < X \le b)$ are mutually exclusive and

$(-\infty < X \le a) \cup (a < X \le b) = (-\infty < X \le b)$

$\therefore P(-\infty < X \le a) + P(a < X \le b) = P(-\infty < X \le b)$

$\Rightarrow F(a) + P(a < X \le b) = F(b)$

$\Rightarrow P(a < X \le b) = F(b) - F(a)$.

③ $F(x)$ is monotonically increasing function i.e.

if $x_2 > x_1 \Rightarrow F(x_2) \geq F(x_1)$

as $0 \le P(x_1 < X \le x_2) = F(x_2) - F(x_1)$ (when $x_2 > x_1$)

$\Rightarrow F(x_2) \geq F(x_1)$

④ $F(x)$ is continuous to the right at every point a

i.e. $\lim_{x \to a^+} F(x) = F(a)$. or, $F(a+0) = F(a)$.

(Proof: Out of scope/syllabus)

⑤ For any real constant a

$F(a) - \lim_{x \to a^-} F(x) = F(a) - F(a-0) = P(X = a)$.

(Proof: Out of scope/syllabus)

⑥ & ⑦ $F(-\infty) = 0$ & $F(\infty) = 1$ where

$F(-\infty) = \lim_{x \to -\infty} F(x)$ & $F(\infty) = \lim_{x \to \infty} F(x)$.

VIII. *The set of points of discontinuity of a distribution function is at most enumerable.* (countable i.e. having a bijection with N or finite).

We know that every monotonic function can have at most a countable set of points of discontinuity. Since every distribution function is monotonic, the property follows.

Remark : (*a*) From properties I-VII we conclude that the distribution function $F(x)$ is a monotonic non-decreasing bounded function such that

(*i*) $F(-\infty)=0$

(*ii*) $F(\infty)=1$

(*iii*) it is continuous to the right at all points

(*iv*) it is discontinuous to the left at every point $x=a$, if $P(X=a)>0$ and the discontinuity being a jump discontinuty, the height of the jump (or saltus) is equal to $P(X=a)$.

(*b*) The converse of the remark (*a*) is also true and so we conclude the following (without proof) :

Any function $F(x)$ with domain $(-\infty, \infty)$ and range a subset of $[0, 1]$ is a distribution function of a random variable with respect to a probability space (S, Δ, P) if and only if $F(x)$ is such that (*i*) $F(-\infty)=0$, (*ii*) $F(\infty)=1$, (*iii*) $F(x)$ is monotonically non-decreasing and bounded, (*iv*) $F(x)$ is continuous to the right at all points, (*v*) $F(x)$ is discontinuous to the left at every point $x=a$, if $P(X=a)>0$.

(*c*) The curve $y=F(x)$ is called the *distribution curve* of the corresponding random variable X. It is evident that the distribution curve lies between $y=0$, $y=1$.

Probability distribution and the concept of probability mass.

If the distribution function $F(x)$ of a random variable X be known, then for any a, b $(a<b)$, the probability of the event $(a<X\leqslant b)$ can be determined. So the distribution function $F(x)$ gives the distribution of probabilites of various events and so we say that $F(x)$ determines the *probability distribution* of the random variable X. Then the problem of determination of the

probability distribution of X is the same as the problem of finding the distribution function $F(x)$ of X.

From the properties of the distribution function proved in § 5.2 (I-VII), it will be possible to make an analogy with 'probability of an event' and 'mass of a particle or of a system of particles.' The aforesaid analogy can be done as follows: We assume that a certain amount of matter is distributed on a given straight line (on which x is measured) in such a way that the total mass of the matter distributed from $-\infty$ up to the point $x=a$ is equal to $F(a)$, where $F(x)$ is the distribution function of the random variable X. Then the property '$F(\infty)=1$' implies immediately that the total mass of matter distributed on the line is 1 unit. The property '$P(a < X \leqslant b) = F(b) - F(a)$' reflects that the probability of the event $(a < X \leqslant b)$ is equal to the mass of the matter distributed on the semi-closed interval $(a, b]$. The relation '$P(X=a)=F(a)-F(a-0)$' shows that the probability of the event $(X=a)$ can be interpreted as the mass of a particle placed at the point $x=a$.

The hypothetical distribution of mass described above is called the *probability mass* and in many situations it will be convenient to think probability in terms of mass by the aforesaid analogy where the probabilty of an event is identified with the mass of a certain amount of matter.

We shall restrict our discussion to two types (unless otherwise stated) of random variables, namely discrete and continuous which will be explained in the following sections.

5.3. Discrete Distribution. Probability Mass Function (p.m.f.)

A random variable X defined on an event space S is said to be discrete if the spectrum of X is at most countable, *i.e.*, if the spectrum is finite or countably infinite. In this case the probability distribution of X will be called a *discrete distribution*.

Let the spectrum of X be $\{ x_i : i=0, \pm 1, \pm 2, \ldots\ldots\ldots\ldots \}$, where $\ldots\ldots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \ldots\ldots$

Let $P(X=x_i)=f_i$, x_i being a spectrum point. A function $f : R \rightarrow [0, 1]$ is defined as follows:

$$f(x)=f_i, \text{ if } x=x_i, \text{ which is a point of the spectrum,}$$
$$=0, \text{ elsewhere.} \qquad (5.3.1)$$

The function f defined above is called the *probability mass function* (p. m. f.) of the random variable X.

The *distribution function* $F(x)$ *of a discrete random variable* X is given by :

$$F(x) = \sum_{x_j \leq x_i} P(X=x_j) = \sum_{j=-\infty}^{i} f_j, \text{ if } x_i \leq x < x_{i+1}$$

$(i=0, \pm 1, \pm 2, \ldots\ldots)$. (5.3.2)

Thus $F(x)$ is a step function which remains constant over every interval in between two consecutive spectrum points, has a jump discontinuity at each spectrum point x_i, the height of the jump at each point being $f_i = P(X=x_i)$. It is continuous to the right but discontinuous to the left at each spectrum point.

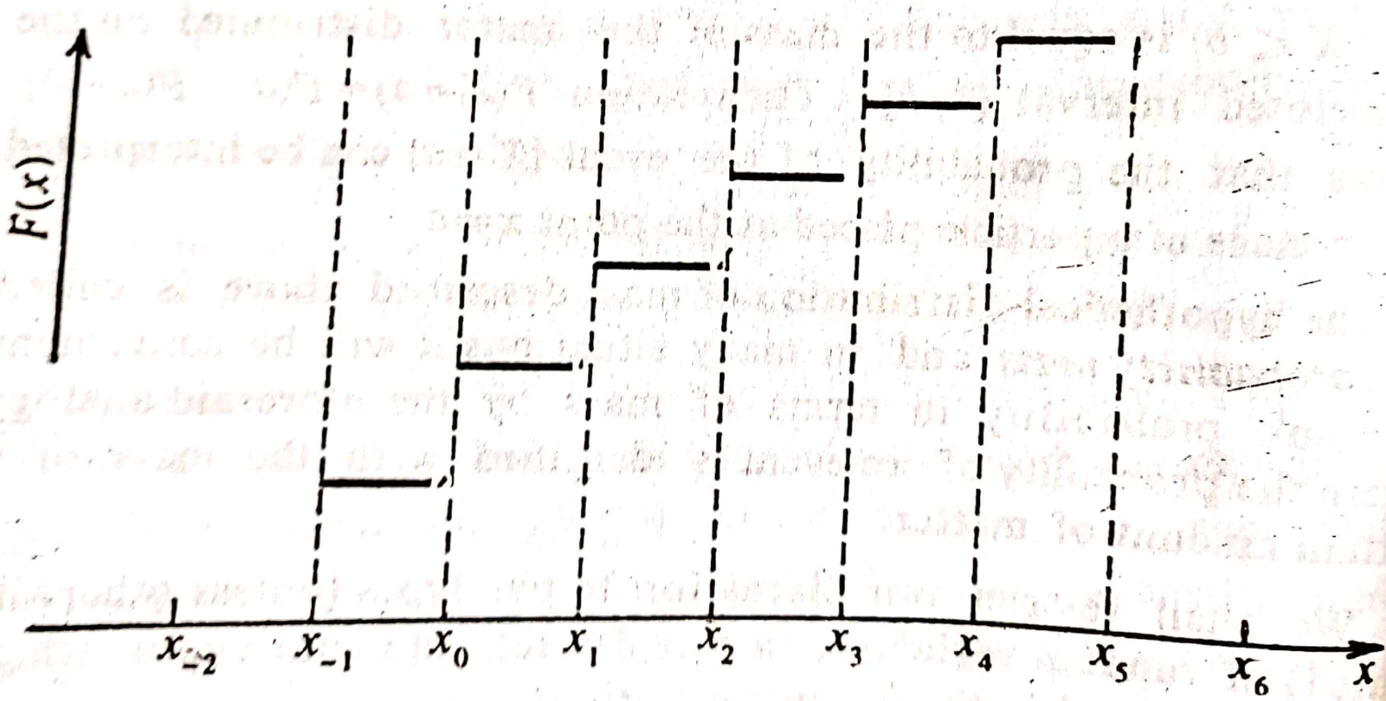


Fig. 5.3.1 Distribution Function of a Discrete Distribution.

Some important results on discrete distributions.

① $\sum_{j=-\infty}^{\infty} f_j = 1$ as $\sum_{j=-\infty}^{\infty} f_j = F(\infty) = 1$

② At each non-spectrum point 'a'.
$P(X=a) = 0$. (Observe the graph & due to the fact $P(X=a) = F(a) - F(a-0)$)

③ $P(a < X \leq b) = \sum_{a < x_i \leq b} f_i$

④ If for X, a discrete random variable with x_i $(i = 0, \pm 1, \pm 2, \cdots)$ as spectrum points. If $P(X=x_i) = f_i$ be given, for $i = 0, \pm 1, \pm 2, \cdots$, the distribution function can be determined & vice-versa as below:

We define the distribution function F as below

$$F(x) = \sum_{j=-\infty}^{i} f_j \ , \text{ for } x_i \leq x < x_{i+1} \ (i = 0, \pm 1, \pm 2, \cdots)$$

& conversely if $F(x)$ as the distribution function of a discrete random variable X of which spectrum is the set $\{x_i : i = 0, \pm 1, \pm 2, \cdots\}$

Then $P(X = x_i)$ denoted as f_i is given by
$f_i = P(X = x_i) = F(x_i) - F(x_i - 0)$ for $i = 0, \pm 1, \pm 2, \cdots$

5.5. Important discrete distributions.

I. *Binomial* (n, p) *Distribution.*

A discrete random variable X having the set $\{0, 1, 2, \ldots\ldots, n\}$ as the spectrum, is said to have *binomial distribution* with parameters n, p if the p. m. f. of X is given by,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x=0, 1, 2, \ldots\ldots, n$$

$$= 0, \text{ elsewhere,}$$

where n is a positive integer and $0 < p < 1$.

We now give one example of binomial distribution from real life situation. Let E_n be the resulting compound experiment arising from n (a positive integer) Bernoulli trials, where $p (0 < p < 1)$ is the probability of success in each trial. If we are interested only in the number of successes, then the event space corresponding to E_n is the finite set $\{0, 1, 2, \ldots\ldots, n\} = S$ (say).

A random variable X is defined on S as follows :

$$X(i) = i \quad \text{where} \quad i \in S.$$

Then X is a discrete random variable where the probability mass function $f(x) = P(X = x)$ is given by

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots n$$

$$= 0, \text{ elsewhere.} \qquad (5.5.1)$$

(5.5.1) shows that X has binomial (n, p) distribution.

II. *Poisson μ Distribution.*

A discrete random variable X having the enumerable set $\{0, 1, 2, \ldots\ldots\}$ as the spectrum, is said to have *Poisson distribution* with parameter $\mu(>0)$, if the p.m.f. is given by

$$f(x)=\frac{e^{-\mu}\mu^x}{x!}, \text{ for } x=0, 1, 2, \ldots\ldots$$
$$=0, \text{ elsewhere.}$$

Let us now give an example of Poisson distribution from real life problems. If X be the random variable denoting the number of telephone calls in a given interval $(0, t)$, satisfying the conditions in a Poisson process (see § 5.11), then X is a discrete variate whose spectrum is the enumerable set $\{0, 1, 2, \ldots\ldots\}$, the corresponding probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!} \text{ for } x=0, 1, 2, \ldots\ldots$$
$$=0, \text{ elsewhere.} \qquad (5.5.2)$$

(λ is the average number of calls per unit time).

(5.5.2) shows that X has Poisson distribution with parameter λt.

5.6. Continuous Random Variable.

Let X be a random variable defined on an event space S. Let $F(x)$ be the distribution function of X. Then the random variable X is said to be continuous if

(*i*) the distribution function $F(x)$ is continuous for all real values of x,

(*ii*) and for any two real numbers a, b $(a < b)$, $\frac{d}{dx}F(x) = F'(x)$ is continuous in $[a, b]$ except for at most a finite number of discontinuities (which may include points of infinite discontinuity) and $\int_a^b F'(x)\,dx$ is convergent.

Alternative definition of continuous variate.

A random variable X defined on the event space S is said to be a continuous random variable if there exists a non-negative real valued function $f(x)$ such that (*i*) $f(x)$ is integrable in $(-\infty, \infty)$ and (*ii*) the distribution function $F(x)$ of X is given by

$F(x) = \int_{-\infty}^{x} f(t)\,dt$ for any real x.

The equivalence of the two definitions will follow from (5.8.3) and note (*d*) of § 5·8. If X is a *continuous random variable*, then the probability distribution of X is called a *continuous distribution*.

5·7. Probability Density Fuuction (p.d.f) of a Continuous Distribution.

In case of a continuous distribution, we denote $F'(x)$ by $f(x)$, where $f(x)$ is called the *probability density function* (p.d.f.) of X, $F(x)$ being the distribution function of X. From definition, the density function is continuous in any finite interval $[a, b]$ except for at most finite number of points of discontinuity. We note that $F'(x)$ may not be defined values of x and consequently $f(x)$ may be undefined at some points.

In the alternative definition of a continuous variate X, the non-negative real valued function $f(x)$ is called a probability density function of X. Here from the relation $F(x) = \int_{-\infty}^{x} f(t)dt$ we get $F'(x) = f(x)$ at a point of continuity x of $f(x)$.

5.6. Continuous Random Variable.

Let X be a random variable defined on an event space S. Let $F(x)$ be the distribution function of X. Then the random variable X is said to be continuous if

(i) the distribution function $F(x)$ is continuous for all real values of x,

(ii) and for any two real numbers a, b $(a < b)$, $\frac{d}{dx}F(x) = F'(x)$ is continuous in $[a, b]$ except for at most a finite number of discontinuities (which may include points of infinite discontinuity) and $\int_a^b F'(x)\,dx$ is convergent.

Alternative definition of continuous variate.

A random variable X defined on the event space S is said to be a continuous random variable if there exists a non-negative real valued function $f(x)$ such that (i) $f(x)$ is integrable in $(-\infty, \infty)$ and (ii) the distribution function $F(x)$ of X is given by

$F(x) = \int_{-\infty}^{x} f(t)\,dt$ for any real x.

The equivalence of the two definitions will follow from (5.8.3) and note (d) of § 5·8. If X is a *continuous random variable*, then the probability distribution of X is called a *continuous distribution.*

5·7. Probability Density Fuuction (p.d.f) of a Continuous Distribution.

In case of a continuous distribution, we denote $F'(x)$ by $f(x)$, where $f(x)$ is called the *probability density function* (p.d.f.) of X, $F(x)$ being the distribution function of X. From definition, the density function is continuous in any finite interval $[a, b]$ except for at most finite number of points of discontinuity. We note that $F'(x)$ may not be defined values of x and consequently $f(x)$ may be undefined at some points.

In the alternative definition of a continuous variate X, the non-negative real valued function $f(x)$ is called a probability density function of X. Here from the relation $F(x) = \int_{-\infty}^{x} f(t)dt$ we get $F'(x) = f(x)$ at a point of continuity x of $f(x)$.

5.8. Some Important Results on the probability density function f and the corresponding distribution function F of a continuous variate X.

I. *$f(x) \geqslant 0$ for all x where $f(x)$ is defined.* (5.8.1)

We know that $F(x)$ is a monotonic increasing function. So $F'(x) \geqslant 0$ whenever $F'(x)$ exists.

$\therefore \quad f(x) \geqslant 0$ for all x, where $f(x)$ is defined.

II. $P(a < X \leqslant b) = \int_a^b f(x)\,dx.$ (5.8.2)

We have $P(a < X \leqslant b) = F(b) - F(a).$

Now $F'(x) = f(x)$ is continuous in $[a, b]$ except for at most a finite number of discontinuities and So, we have $\int_a^b F'(x)\,dx$ is convergent.

Then, $\int_a^b f(x)\,dx = F(b) - F(a) = P(a < X \leqslant b).$

III. $F(x) = \int_{-\infty}^{x} f(t)\,dt.$ (5.8.3)

We have, by (5.8.2),

$$P(a < X \leq x) = \int_a^x f(t)\,dt$$

$$\text{or,} \quad F(x) - F(a) = \int_a^x f(t)\,dt.$$

Proceeding to the limit $a \to -\infty$, we get

$$F(x) - \underset{a \to -\infty}{Lt} F(a) = \underset{a \to -\infty}{Lt} \int_a^x f(t)\,dt$$

$$\text{or,} \quad F(x) - F(-\infty) = \int_{-\infty}^{x} f(t)\,dt$$

$$\text{or,} \quad F(x) = \int_{-\infty}^{x} f(t)\,dt, \text{ since } F(-\infty) = 0.$$

IV. $\int_{-\infty}^{\infty} f(x)\,dx = 1.$ (5.8.4)

We have, by (5.8.3),

$$\int_{-\infty}^{x} f(t)\,dt = F(x).$$

$\therefore$ proceeding to the limit $x \to \infty$,

$$\underset{x \to \infty}{Lt} \int_{-\infty}^{\infty} f(t)\, dt = \underset{x \to \infty}{Lt} F(x) = F(\infty) = 1$$

or, $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

V. $P(X=a)=0$, *where a is a given constant.* (5.8.5)

We have $F(a) - F(a-0) = P(X=a)$.

Now X being a continuous random variable, $F(x)$ is continuous for all x. Hence $F(x)$ is continuous at $x=a$.

$$\therefore \quad \underset{x \to a-0}{Lt} F(x) = F(a),$$

i.e., $F(a-0) = F(a)$.

$$\therefore \quad P(X=a) = F(a) - F(a-0) = 0,$$

i.e., $P(X=a)=0$ for any real constant a.

Note : (*a*) We see that the distribution function of a continuous random variable X is completely determined by the corresponding probability density function $f(x)$, using (5.8.3). So the probability distribution of a continuous random variable is completely determined by the corresponding density function $f(x)$.

(*b*) We observe that the probability density function defined in the two ways mentioned in § 5·7 may differ at some points but they will determine the same distribution function $F(x)$ of a continuous random variable X. Further we observe that if the values of the p.d.f. (in any definition) be altered at finite number of points or if the p.d.f. be defined arbitrarily at finite number of points where it is undefined, then the corresponding distribution function $F(x)$ is not altered.

(*c*) We know that $P(O)=0$. If, however $P(A)=0$, we cannot conclude that A is an impossible event. In this case, we say that A is *stochastically impossible*. We now give an example to show that 'an event may be stochastically impossible but not impossible.'

Let E be the random experiment of selecting a number at random from the open interval (0, 9). Let X be the random variable denoting the number chosen. Then an event '$X=6$' is not an impossible event. But it can be shown that X is a continuous random variable and so $P(X=6)=0$. So the event '$X=6$' is a stochastically impossible event but not an impossible event.

(*d*) *Every non-negative real valued piecewise continuous function* $f(x)$ *that is integrable in* $(-\infty, \infty)$ *and for which* $\int_{-\infty}^{\infty} f(x)\, dx=1$, *is the probability density function of a continuous distribution.*

& (e) $f(x)\,dx = P(x < X \leq x+dx)$. (5.8.6)

It is sufficient to show that there exists a distribution function $F(x)$ corresponding to $f(x)$.

We define a function F given by $\int_{-\infty}^{x} f(t)\, dt=F(x)$, (5.8.7)

Density curve: The graphical representation of $y=f(x)$ is called the probability *density curve* of the corresponding continuous distribution.

VI. *Probability Differential.*

Let X be a continuous random variable and $\delta x > 0$.

Then $P(x < X \leq x+\delta x) = F(x+\delta x) - F(x) = \delta x F'(\xi)$,

$$\xi = x + \theta\delta x,\ 0 < \theta < 1,$$

by Lagrange's Mean Value Theorem of Differential Calculus.

$$\therefore \quad \underset{\delta x \to 0}{Lt} \frac{P(x < X \leq x+\delta x)}{\delta x} = \underset{\delta x \to 0}{Lt} F'(\xi) = F'(x),$$

if x is a point of continuity of $F'(x) = f(x)$.

$$\therefore \quad f(x) = \underset{\delta x \to 0}{Lt} \frac{P(x < X \leq x+\delta x)}{\delta x}$$

$$= \underset{dx \to 0}{Lt} \frac{P(x < X \leq x+dx)}{dx},$$

since $\delta x = dx$, the differential of the variable x.

Henceforth we shall write $f(x)\, dx$ for $P(x < X \leq x+dx)$, which will actually mean $\underset{dx \to 0}{Lt} \frac{P(x < X \leq x+dx)}{dx} = f(x)$.

The expression $P(x < X \leqslant x+dx)$ will always be used in the above limiting sense and so there will be no ambiguity throughout our discussion. The expression $P(x < X \leqslant x+dx)$ which is taken to be equal to $f(x)dx$,

i.e., $f(x)\, dx = P(x < X \leqslant x+dx)$ (5.8.8)

is called the *probability differential* of the continuous random variable X.

We now discuss some important continuous distributions.

5.9. Important Continuous Distributions.

I. *Uniform or Rectangular Distribution.*

A continuous random variable X, is said to follow a uniform distribution, if its probability density function (p.d.f) is given by

$$f(x) = \frac{1}{b-a}, \; a < x < b$$
$$= 0, \text{ elsewhere,} \tag{5.9.1}$$

where a and b are the two parameters of the distribution.

We note that $f(x) \geqslant 0$ for all x.

Also $\int_{-\infty}^{\infty} f(x)\, dx = \int_a^b \frac{dx}{b-a} = 1.$

The distribution function $F(x)$ of X is given by

$$F(x) = 0, \; -\infty < x < a$$
$$= \frac{x-a}{b-a}, \; a \leqslant x \leqslant b$$
$$= 1, \; b < x < \infty. \tag{5.9.2}$$

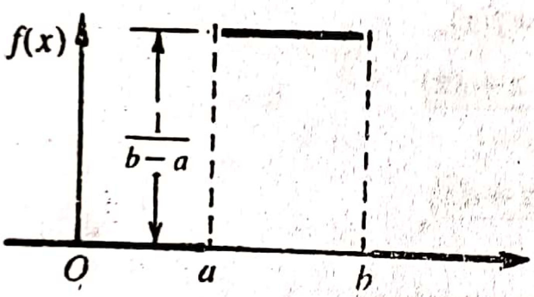


Fig. 5.9.1
Rectangular Density Curve

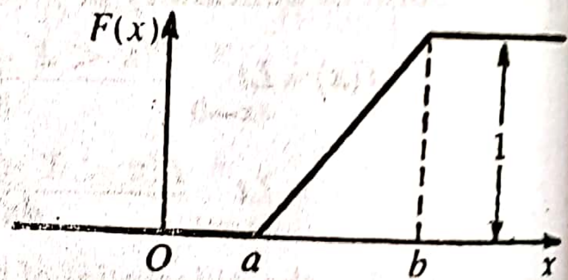


Fig. 5.9.2
Rectangular Distribution Curve

Note : The rectangular distribution gives a useful model for random experiment like 'a point is chosen at random in a given interval'. In this case, we are actually thinking of a random variable X such that the probability of the event 'X lying in any sub-interval' is proportional to the length of the sub-interval, *i.e.*, X is uniformly distributed in the given interval.

II. *Normal* (m, σ) *Distribution.*

A continuous random variable X, having $(-\infty, \infty)$ as the spectrum, is said to follow a *normal distribution* if its probability density function $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty \tag{5.9.3}$$

where $\sigma > 0$.

It is a probability distribution with two parameters m, σ and is denoted by $N(m, \sigma)$.

In particular if $m=0$, $\sigma=1$, we say that the corresponding random variable X is a *Standard Normal Variate*.

Since $\frac{1}{\sqrt{2\pi}\,\sigma} > 0$ and $e^{-\frac{(x-m)^2}{2\sigma^2}}$ is non-negative, for all values of x, hence $f(x) > 0$ for all x.

Again $\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\int_{P}^{Q} e^{\frac{-(x-m)^2}{2\sigma^2}}\,dx$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\int_{\frac{P-m}{\sqrt{2}\sigma}}^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,.\,\sqrt{2}\sigma\,dy \quad \text{where } \frac{x-m}{\sqrt{2}\sigma}=y$$

$$=\frac{1}{\sqrt{\pi}}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\left[\int_{\frac{P-m}{\sqrt{2}\sigma}}^{0} e^{-y^2}\,dy+\int_{0}^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy\right]$$

$$=\frac{1}{\sqrt{\pi}}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\left[-\int_{\frac{-P+m}{\sqrt{2}\sigma}}^{0} e^{-t^2}\,dt+\int_{0}^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy\right]$$

where in the first integral we put $y=-t$

$$=\frac{1}{\sqrt{\pi}}\ \underset{P\to-\infty}{Lt}\int_{0}^{\frac{-P+m}{\sqrt{2}\sigma}} e^{-t^2}\,dt+\frac{1}{\sqrt{\pi}}\ \underset{Q\to\infty}{Lt}\int_{0}^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy$$

$$=\frac{1}{\sqrt{\pi}}\int_{0}^{\infty} e^{-t^2}\,dt+\frac{1}{\sqrt{\pi}}\int_{0}^{\infty} e^{-y^2}\,dy,$$

since the integrals are convergent

$$=\frac{2}{\sqrt{\pi}}\int_{0}^{\infty} e^{-y^2}\,dy=\frac{2}{\sqrt{\pi}}\left[\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right]=\frac{2}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{2}=1.$$

The distribution function of a normal distribution is given by

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx. \tag{5.9.4}$$

If X is a standard normal variate (*i.e.*, $m=0$, $\sigma=1$), the corresponding density and distribution functions are given by

$$\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \text{ and } \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt. \tag{5.9.5}$$

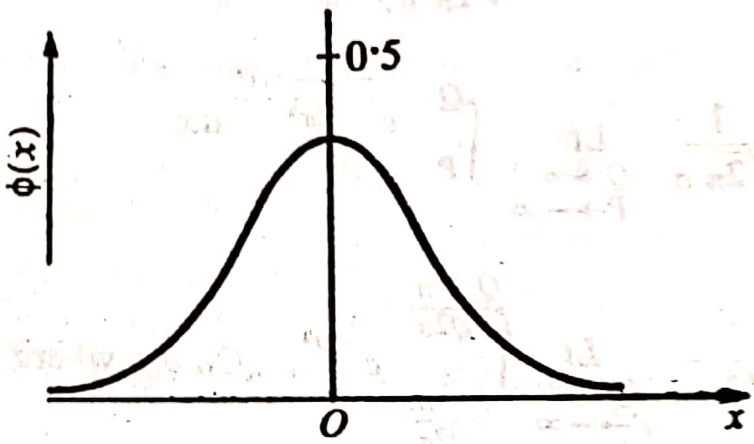


Fig. 5.9.3 Standard Normal Density Curve

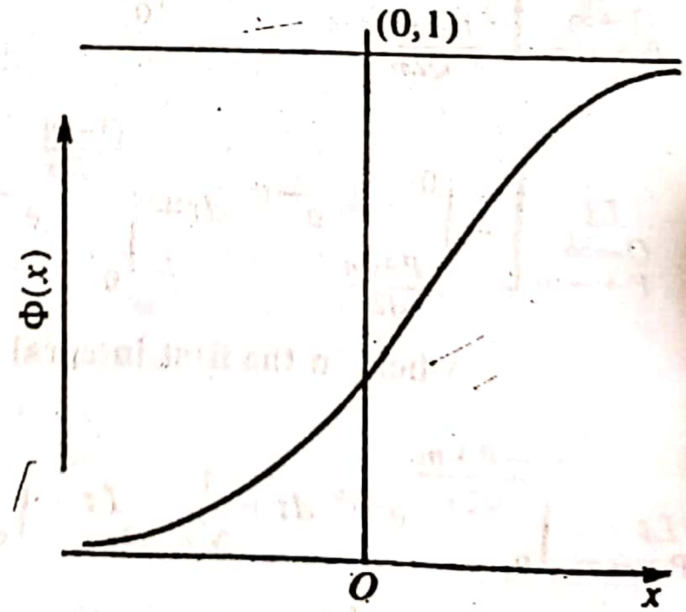


Fig. 5.9.4 Standard Normal Distribution Curve

One of the most important distributions in the theory of probability and statistics is the normal distribution and in the foregoing chapters we study this distribution in detail.

5.10. Distinction between Discrete and Continuous Random Variables.

(*a*) A random variable X is discrete when its spectrum is at most an enumerable set, *i.e.*, either finite or countably infinite, whereas in the case of a continuous random variable, the spectrum is usually an interval or union of some intervals.

(*b*) The distribution function of a discrete random variable is a step function, whereas in the case of a continuous random variable, the distribution function $F(x)$ is continuous for all x and in any bounded interval $[a, b]$ $F'(x)$ is continuous except for at most a finite number of points of discontinuity.

(*c*) In the case of a discrete random variable $P(X=a)=0$ if a is not a spectrum point, while in the case of a continuous random variable $P(X=a)=0$ for any real number a.

(*d*) The random variable denoting the number of telephone calls in a given trunk line in a given interval of time is an example of a discrete random variable (see Poisson Process). The random variable denoting the number chosen at random from a given interval, say (4, 7), is an example of a continuous random variable.

Transformation of continuous random variable

Theorem 3. Let X be an RV of the continuous type with PDF f. Let $y = g(x)$ be differentiable for all x and either $g'(x) > 0$ for all x or $g'(x) < 0$ for all x. Then $Y = g(X)$ is also an RV of the continuous type with PDF given by

$$(2) \qquad h(y) = \begin{cases} f[g^{-1}(y)] \left| \dfrac{d}{dy} g^{-1}(y) \right|, & \alpha < y < \beta, \\ 0, & \text{otherwise}, \end{cases}$$

where $\alpha = \min\{g(-\infty), g(+\infty)\}$ and $\beta = \max\{g(-\infty), g(+\infty)\}$.

Proof. If g is differentiable for all x and $g'(x) > 0$ for all x, then g is continuous and strictly increasing, the limits α, β exist (may be infinite), and the inverse function $x = g^{-1}(y)$ exists, is strictly increasing, and is differentiable. The DF of Y for $\alpha < y < \beta$ is given by

$$P\{Y \leq y\} = P\{X \leq g^{-1}(y)\}. = F(g^{-1}(y))$$

The PDF of g is obtained on differentiation. We have

$$\begin{aligned} h(y) &= \frac{d}{dy} P\{Y \leq y\} \\ &= f[g^{-1}(y)] \frac{d}{dy} g^{-1}(y). \end{aligned}$$

Similarly, if $g' < 0$, then g is strictly decreasing and we have

$$\begin{aligned} P\{Y \leq y\} &= P\{X \geq g^{-1}(y)\} \\ &= 1 - P\{X \leq g^{-1}(y)\} \qquad (X \text{ is a continuous RV}) \end{aligned}$$

so that

$$h(y) = -f[g^{-1}(y)] \cdot \frac{d}{dy} g^{-1}(y).$$

Since g and g^{-1} are both strictly decreasing, $(d/dy)\, g^{-1}(y)$ is negative and (2) follows.

Note: RV ≡ random variable
PDF ≡ Probability density function.

Note that

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{dg(x)/dx}\bigg|_{x=g^{-1}(y)},$$

so that (2) may be rewritten as

$$(3) \qquad h(y) = \frac{f(x)}{|dg(x)/dx|}\bigg|_{x=g^{-1}(y)}, \qquad \alpha < y < \beta.$$

Remark 1. The key to computation of the induced distribution of $Y = g(X)$ from the distribution of X is (1). If the conditions of Theorem 3 are satisfied, we are able to identify the set $\{X \in g^{-1}(-\infty, y]\}$ as $\{X \leq g^{-1}(y)\}$ or $\{X \geq g^{-1}(y)\}$, according to whether g is increasing or decreasing. In practice, Theorem 3 is quite useful, but whenever the conditions are violated, one should return to (1) to compute the induced distribution. This is the case, for example, in Examples 7 and 8 and Theorem 4 below.

Remark 2. If the PDF f of X vanishes outside an interval $[a, b]$ of finite length, we need only to assume that g is differentiable in (a, b), and either $g'(x) > 0$ or $g'(x) < 0$ throughout the interval. Then we take

$$\alpha = \min\{g(a), g(b)\} \quad \text{and} \quad \beta = \max\{g(a), g(b)\}$$

in Theorem 3.

Example 5. Let X have the density $f(x) = 1$, $0 < x < 1$, and $= 0$ otherwise. Let $Y = e^X$. Then $X = \log Y$, and we have

$$h(y) = \left|\frac{1}{y}\right| \cdot 1, \qquad 0 < \log y < 1,$$

that is,

$$h(y) = \begin{cases} \dfrac{1}{y}, & 1 < y < e, \\ 0, & \text{otherwise.} \end{cases}$$

If $y = -2\log x$, then $x = e^{-y/2}$ and

$$\begin{aligned} h(y) &= |-\tfrac{1}{2}e^{-y/2}| \cdot 1, & 0 < e^{-y/2} < 1, \\ &= \begin{cases} \tfrac{1}{2}e^{-y/2}, \\ 0, \end{cases} & \begin{matrix} 0 < y < \infty, \\ \text{otherwise.} \end{matrix} \end{aligned}$$

Example 6. Let X be a nonnegative RV of the continuous type with PDF f, and let $\alpha > 0$. Let $Y = X^{\alpha}$. Then

$$P\{X^{\alpha} \le y\} = \begin{cases} P\{X \le y^{1/\alpha}\} & \text{if } y \ge 0, \\ 0 & \text{if } y < 0. \end{cases}$$

The PDF of Y is given by

$$\begin{aligned} h(y) &= f(y^{1/\alpha}) \left| \frac{d}{dy} y^{1/\alpha} \right| \\ &= \begin{cases} \dfrac{1}{\alpha} y^{1/\alpha - 1} f(y^{1/\alpha}), & y > 0, \\ 0, & y \le 0. \end{cases} \end{aligned}$$

Example 7. Let X be an RV with PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

Let $Y = X^2$. In this case, $g'(x) = 2x$, which is > 0 for $x > 0$, and < 0 for $x < 0$, so that the conditions of Theorem 3 are not satisfied. But for $y > 0$,

$$\begin{aligned} P\{Y \le y\} &= P\{-\sqrt{y} \le X \le \sqrt{y}\} \\ &= F(\sqrt{y}) - F(-\sqrt{y}), \end{aligned}$$

where F is the DF of X. Thus the PDF of Y is given by

$$h(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})], & y > 0, \\ 0, & y \le 0. \end{cases}$$

Thus

$$h(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi\, y}} e^{-y/2}, & 0 < y, \\ 0, & y \le 0. \end{cases}$$

Example 8. Let X be an RV with PDF

$$f(x) = \begin{cases} \dfrac{2x}{\pi^2}, & 0 < x < \pi, \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y = \sin X$. In this case $g'(x) = \cos x > 0$ for x in $(0, \pi/2)$ and < 0 for x in $(\pi/2, \pi)$, so that the conditions of Theorem 3 are not satisfied. To compute the PDF

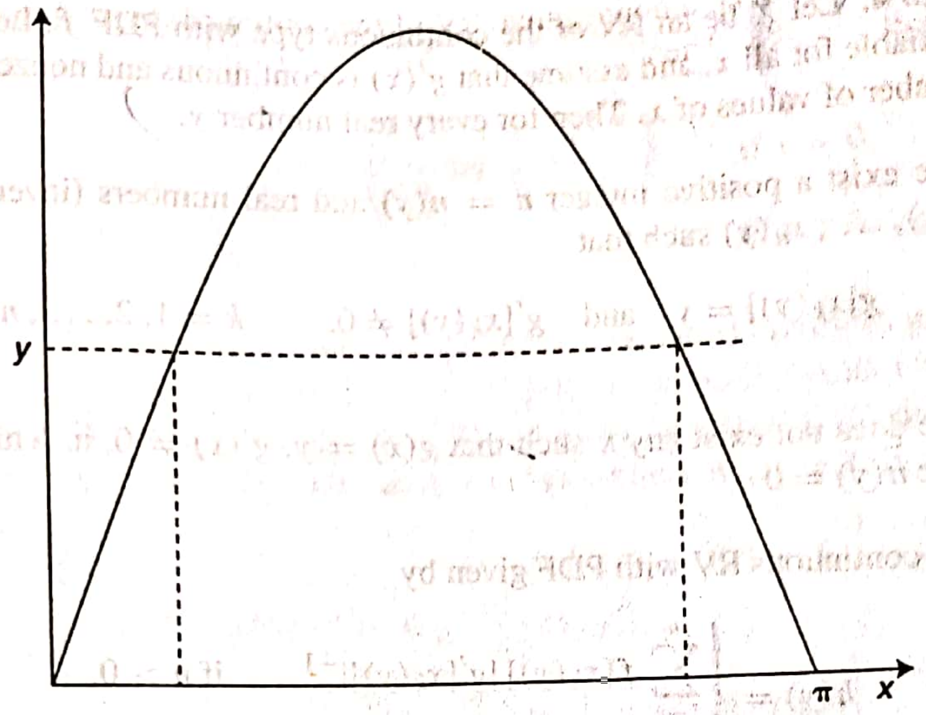


Fig. 1. $y = \sin x,\ 0 \le x \le \pi.$

of Y, we return to (1) and see that (Fig. 1) the DF of Y is given by

$$\begin{aligned} P\{Y \le y\} &= P\{\sin X \le y\}, \qquad 0 < y < 1, \\ &= P\{(0 \le X \le x_1) \cup (x_2 \le X \le \pi)\}, \end{aligned}$$

where $x_1 = \sin^{-1} y$ and $x_2 = \pi - \sin^{-1} y$. Thus

$$\begin{aligned} P\{Y \le y\} &= \int_0^{x_1} f(x)\,dx + \int_{x_2}^{\pi} f(x)\,dx \\ &= \left(\frac{x_1}{\pi}\right)^2 + 1 - \left(\frac{x_2}{\pi}\right)^2, \end{aligned}$$

and the PDF of Y is given by

$$\begin{aligned} h(y) &= \frac{d}{dy}\left(\frac{\sin^{-1} y}{\pi}\right)^2 + \frac{d}{dy}\left[1 - \left(\frac{\pi - \sin^{-1} y}{\pi}\right)^2\right] \\ &= \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

5.13. Mixed Distribution.

So far we have discussed distributions which are either discrete or continuous. But there are distributions which are neither discrete nor continuous. In fact, there are probability distributions where the corresponding distribution is partly discrete and partly continuous. Such a distribution is called a mixed distribution. We give below a formal definition of a mixed distribution.

A distribution is called a *mixed distribution,* if the corresponding distribution function $F(x)$ can be expressed as a convex combination of the form

$$F(x)=cF_1(x)+(1-c)F_2(x) \tag{5.13.1}$$

where $F_1(x)$ is the distribution function of a discrete random variable and $F_2(x)$ is that of a continuous random variable and $0 < c < 1$.

We complete the discussion by giving the following example of a mixed distribution.

Let X be a random variable with distribution function $F(x)$ given by

$$F(x)=\begin{cases} 0, & x<0 \\ \frac{x+1}{2}, & 0 \leq x < 1 \\ 1, & 1 \leq x. \end{cases}$$

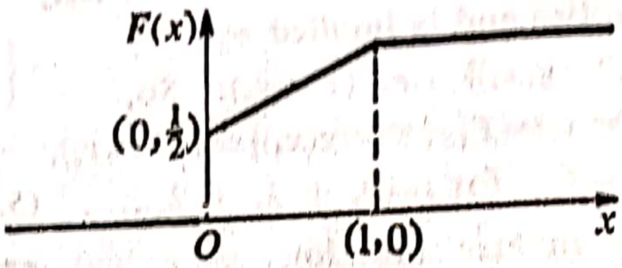


Fig. 5.13.1 Distribution curve of a mixed distribution.

From fig. 5.13.1, we see that $F(x)$ has a jump discontinuity at $x=0$. In fact, $F(x)$ is not always continuous, nor is it a step function. Accordingly, the corresponding distribution is a mixed distribution.

We can write

$$F(x)=\tfrac{1}{2}F_1(x)+\tfrac{1}{2}F_2(x),$$

where $$F_1(x)=\begin{cases} 0, & x<0 \\ 1, & x \geq 0 \end{cases}$$

and $$F_2(x)=\begin{cases} 0, & x<0 \\ x, & 0<x<1 \\ 1, & x \geq 1. \end{cases}$$

$F_1(x)$ and $F_2(x)$ are the distribution functions of a discrete and a continuous distribution respectively. The probability density function $f(x)$ corresponding to $F_2(x)$ is given by

$$f(x)=\begin{cases} 1, & 0<x<1 \\ 0, & \text{if } -\infty<x<0, \text{ and } 1<x<\infty. \end{cases}$$

We observe that $f(x)$ is undefined at $x=0$ and at $x=1$.

In fact $L\,F_2'(0)=0$, $R\,F_2'(0)=1$, etc.

5·14. Illustrative Examples.

Ex. 1. *Five balls are drawn from an urn containing 4 white and 6 black balls. Find the probability distribution of the number of white balls drawn without replacement.*

Let X be the random variable denoting the number of white balls drawn from the urn. Then the spectrum of X is the set $\{0, 1, 2, 3, 4\}$.

Now $P(X=0)=\dfrac{{}^6P_5}{{}^{10}P_5}=\dfrac{1}{42}$, $P(X=1)=\dfrac{{}^4C_1\times{}^6C_4\times\lfloor 5}{{}^{10}P_5}=\dfrac{5}{21}$,

$$P(X=2)=\frac{{}^4C_2\times{}^6C_3\times\lfloor 5}{{}^{10}P_5}=\frac{10}{21},$$

$$P(X=3)=\frac{{}^4C_3\times{}^6C_2\times\lfloor 5}{{}^{10}P_5}=\frac{5}{21},$$

$$P(X=4)=\frac{{}^4C_4\times{}^6C_1\times\lfloor 5}{{}^{10}P_5}=\frac{1}{42}.$$

Hence the required distribution at X is given by the spectrum $\{0, 1, 2, 3, 4\}$ with

$P(X=0)=\frac{1}{42}$, $P(X=1)=\frac{5}{21}$, $P(X=2)=\frac{10}{21}$, $P(X=3)=\frac{5}{21}$, and $P(X=4)=\frac{1}{42}$.

Ex. 2. *Consider the random experiment of tossing a fair coin till a head appears for the first time. Let X be the number of tosses required. Find the distribution of X.*

The spectrum of X is the set $\{1, 2, 3, \ldots\ldots\}$.

Now $P(X=1)=P\{H\}=\frac{1}{2}$, $P(X=2)=P\{(T, H)\}=(\frac{1}{2})^2$,

$P(X=3)=P\{(T, T, H)\}=(\frac{1}{2})^3, \ldots\ldots,$

$$P(X=n)=P\{(\underbrace{T, T, \ldots\ldots T}_{(n-1)\text{ times}}, H)\}=(\tfrac{1}{2})^n$$

and so on, where, 'H' denotes the outcome head in the 'first toss', (T, H) denotes the outcome 'Tail in the first toss and head in the second toss', etc. Thus the required distribution is given by $X=i$, $i=1, 2, \ldots\ldots$, with

$$P(X=i)=(\tfrac{1}{2})^i.$$

Ex. 3. *Show that the function* $|x|$ *in* $(-1, 1)$ *and zero elsewhere is a possible density function, and find the corresponding distribution function.*

Let $f(x)=\begin{cases} |x|, & -1<x<1 \\ 0, & \text{elsewhere.} \end{cases}$

We see that $f(x) \geqslant 0$ for every x.

Also, $\int_{-\infty}^{\infty} f(x)\,dx=\int_{-\infty}^{-1} f(x)\,dx+\int_{-1}^{1} f(x)\,dx+\int_{1}^{\infty} f(x)\,dx$

$$=0+\int_{-1}^{1} |x|\,dx+0$$

$$=2\int_{0}^{1} x\,dx, \text{ since } |x| \text{ is an even function and } |x|=x \text{ for every } x \in (0,1)$$

$$=1.$$

Hence, $f(x)$ is a possible probability density function of some distribution.

Now let $F(x)$ be the corresponding distribution function.

If $-\infty<x\leqslant -1$, $F(x)=\int_{-\infty}^{x} f(t)\,dt=0.$

If $-1<x<0$, $F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{x} f(t)\,dt$

$$=0+\int_{-1}^{x} f(t)\,dt=\int_{-1}^{x} (-t)\,dt=\frac{1}{2}-\frac{x^2}{2}.$$

If $0\leqslant x<1$, $F(t)=\int_{-\infty}^{x} f(t)\,dt$

$$=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{0} f(t)\,dt+\int_{0}^{x} f(t)\,dt$$

$$=0-\int_{-1}^{0} t\,dt+\int_{0}^{x} t\,dt$$

$$=\frac{1}{2}+\frac{x^2}{2}.$$

If $1\leqslant x<\infty$, $F(x)=\int_{-\infty}^{x} f(t)\,dt$

$$=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{1} f(t)\,dt+\int_{1}^{x} f(t)\,dt$$

$$=0+1+0=1.$$

Hence the distribution function is given by

$$F(x)=\begin{cases} 0 & , -\infty < x \leqslant -1 \\ \frac{1}{2}-\frac{x^2}{2}, & -1 < x < 0 \\ \frac{1}{2}+\frac{x^2}{2}, & 0 \leqslant x < 1 \\ 1 & , 1 \leqslant x < \infty. \end{cases}$$

Ex. 4. *Can the following be probability mass functions ?*

(a) $f(x)=\begin{cases} 2 \text{ for } x=\frac{1}{2} \\ 1 \text{ for } x=\frac{1}{4} \\ -1 \text{ for } x=\frac{3}{4} \\ 0, \text{ elsewhere.} \end{cases}$ (b) $f(x)=\begin{cases} \frac{1}{8} \text{ for } x=1 \\ \frac{2}{8} \text{ for } x=2 \\ \frac{3}{8} \text{ for } x=3 \\ 0, \text{ elsewhere.} \end{cases}$

(c) $f(x)=\begin{cases} 0{\cdot}1 \text{ for } x=-5 \\ 0{\cdot}5 \text{ for } x=-1 \\ 0{\cdot}2 \text{ for } x=0 \\ 0{\cdot}2 \text{ for } x=1 \\ 0, \text{ elsewhere.} \end{cases}$

(a) Since $f(\frac{3}{4})=-1<0$, $f(x)$ is not a probability mass function.

(b) Although $f(x) \geqslant 0$ for every mass point, $\sum_x f(x) \neq 1$.

Hence $f(x)$ is not a probability mass function.

(c) $f(x) \geqslant 0$ for every spectrum point and $\sum_x f(x)=1$, hence $f(x)$ is a probability mass function of a distribution.

Ex. 5. *Evaluate the distribution function of the following distribution : Spectrum of the random variable X is* $\{-1, 0, 2, 3\}$ *with*

$$P(X=-1)=\tfrac{1}{7},\ P(X=0)=\tfrac{2}{7},\ P(X=2)=\tfrac{3}{7},\ P(X=3)=\tfrac{1}{7}.$$

Let $F(x)$ be the distribution function.

If $-\infty < x < -1$, $F(x)=0$.

If $-1 \leqslant x < 0$, $F(x)=P(X=-1)=\frac{1}{7}$.

If $0 \leqslant x < 2$, $F(x)=P(X=-1)+P(X=0)=\frac{1}{7}+\frac{2}{7}=\frac{3}{7}$.

If $2 \leqslant x < 3$, $F(x)=P(X=-1)+P(X=0)+P(X=2)$

$$=\tfrac{1}{7}+\tfrac{2}{7}+\tfrac{3}{7}=\tfrac{6}{7}.$$

If $3 \leqslant x < \infty$, $F(x)=P(X=-1)+P(X=0)+P(X=2)+P(X=3)$

$$=\tfrac{1}{7}+\tfrac{2}{7}+\tfrac{3}{7}+\tfrac{1}{7}$$

$$=1.$$

Ex. 6. *Let $F(x)$ be the distribution function of a random variable X. Prove that*

(i) $P(a \leqslant X < b)=F(b-0)-F(a-0)$,

(ii) $P(a \leqslant X \leqslant b)=F(b)-F(a-0)$.

(i) The event $(a \leq X<b)$ can be expressed as $(a<X<b)+(X=a)$.

$$\therefore \quad P(a \leqslant X < b)=P(a<X<b)+P(X=a), \qquad (5.14.1)$$

where we note that $(a<X<b)$, $(X=a)$ are mutually exclusive events.

Again we can write

$$(a<X \leqslant b)=(a<X<b)+(X=b).$$

So $P(a<X \leq b)=P(a<X<b)+P(X=b)$.

$$P(a<X<b)=P(a<X \leqslant b)-P(X=b).$$

Hence, by (5.14.1) we get

$$P(a \leqslant X < b)=P(a<X \leqslant b)-P(X=b)+P(X=a)$$

$$=F(b)-F(a)-F(b)+F(b-0)+F(a)-F(a-0)$$

$$=F(b-0)-F(a-0).$$

(ii) We have

$$(a \leqslant X \leq b)=(a<X \leq b)+(X=a),$$

where $(a<X \leq b)$, $(X=a)$ are two mutually exclusive events.

So $P(a \leqslant X \leqslant b)=P(a<X \leq b)+P(X=a)$

$$=F(b)-F(a)+F(a)-F(a-0)$$

$$=F(b)-F(a-0).$$

Ex. 7. *Can the following function be a distribution function?*

$$F(x)=\begin{cases} 0, & -\infty<x<0 \\ \frac{1}{5}, & 0 \leqslant x < 1 \\ \frac{3}{5}, & 1 \leqslant x < 3 \\ 1, & 3 \leqslant x < \infty. \end{cases}$$

If so, find the spectrum and probability mass function.

It is clear that $F(x)$ is monotonically non-decreasing and non-negative and $F(\infty)=1$, $F(-\infty)=0$. $F(x)$ is a step function, discontinuous to the left of the three step points 0, 1, 3 and continuous to the right everywhere. Hence, $F(x)$ is a possible distribution function of a discrete random variable X. The spectrum of

X is $\{0, 1, 3\}$ with

$$P(X=0)=F(0)-F(0-0)=\tfrac{1}{8},$$
$$P(X=1)=F(1)-F(1-0)=\tfrac{3}{8}-\tfrac{1}{8}=\tfrac{2}{8},$$
$$P(X=3)=F(3)-F(3-0)=1-\tfrac{3}{8}=\tfrac{5}{8},$$

which give the probability masses at the spectrum points and these probability masses determine the required probability mass function.

Ex. 8. *Determine the value of the constant C such that $f(x)$ defined by*

$$f(x)=\begin{cases} Cx(1-x), & 0<x<1 \\ 0, & \textit{elsewhere} \end{cases}$$

is a probability density function. Find the corresponding distribution function and $P(X > \frac{1}{8})$.

In order that $f(x)$ is a possible probability density function, we must have

$$\int_{-\infty}^{\infty} f(x)\,dx=1$$

$$\text{i.e., } C\int_0^1 x(1-x)\,dx=1$$

i.e., $C=6$.

Let $F(x)$ be the corresponding distribution function.

In $-\infty<x<0$, $F(x)=0$,

in $0 \le x \le 1$, $F(x)=6\int_0^x t(1-t)\,dt=3x^2-2x^3$,

in $1<x<\infty$, $F(x)=6\int_0^1 t(1-t)\,dt=1$.

$$\therefore\quad F(x)=\begin{cases} 0, & -\infty<x<0 \\ 3x^2-2x^3, & 0 \leqslant x \leqslant 1 \\ 1, & 1<x<\infty. \end{cases}$$

Ex. 17. *The probability of a product produced by a machine to be defective is 0·01. If 30 products are taken at random, find the probability that exactly 2 will be defective. Approximate by Poisson distribution and evaluate the error in the approximation.*

As in Ex. 16, required probability $=\binom{30}{2}(\cdot 01)^2(\cdot 99)^{28}=0\cdot 0276$.

Since the probability of success is small, we approximate by Poisson distribution, the parameter of the distribution being $\mu=np=30\times\cdot 01=0\cdot 3$.

Hence the probability of getting exactly 2 defective

$$=\frac{\mu^2}{2\,!}\,e^{-\mu}=\frac{(0\cdot 3)^2}{2\,!}\,e^{-0\cdot 3}=0\cdot 03337.$$

$\therefore$ the error in the approximation $=0\cdot 03337-0\cdot 03276=0\cdot 00061$.

Ex. 19. *If there is a war every 15 years on the average, then find the probability that there will be no war in 25 years.*

λ = number of changes per unit of time on the average $=\frac{1}{15}$. Let X be the random variable denoting the number of wars in the interval (0, 25), when the unit of time is one year, then X is Poisson distributed with parameter $\mu=\lambda t=\frac{1}{15}\times 25=\frac{5}{3}$.

$\therefore$ probability of no war in the given interval of time

$$=P(X=0)=\frac{e^{-\mu}\mu^{0}}{0\,!}=e^{-\frac{5}{3}}.$$

Expected value of a RV (Random Variable) is known as expectation.

Case 1 <u>Discrete Random Variable:-</u> Let X be a discrete random variable with spectrum set $\{\dots, x_{-3}, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots\}$ then its expectation is given by

$$\mu = E(X) = \sum_{r=-\infty}^{\infty} x_r \, P(X = x_r)$$

Ex: X ⇝ Number of heads in two successive coin toss.
Probability mass function of X is given by

i	-1	0	1
x_i	0	1	2
$P(X=x_i)$	1/4	1/2	1/4
Event $(X=x_i)$	{TT}	{HT, TH}	{TT}

$\therefore$ Expectation $= 0\times\frac{1}{4} + 1\times\frac{1}{2} + 2\times\frac{1}{4}$
$= 1.$

Case 2 <u>Continuous Random Variable:-</u> If X is a continuous random variable with pdf $f(x)$, then the expected value (expectation/ mean) of X is given by

$$\mu = \mu_X = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx.$$

Ex: Let pdf of a cont. random variable X be

$$f(x) = \begin{cases} x & 0 \le x \le 1 \\ 2-x & 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore E(X) = \int_0^1 x \cdot x\, dx + \int_1^2 x\cdot(2-x)\, dx = \frac{1}{3} + \frac{2}{3} = 1 \text{ (Ans)}$$

Case 3 <u>Mixed Random Variable:-</u>

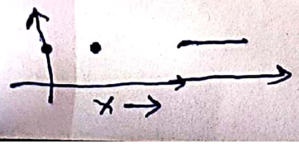
Expectation = Expectation for discrete part + Expectation for continuous part

Ex:- Let X is a mixed random variable with following probability function

$P(X=0) = 1/4$, $P(X=1) = 1/4$, $f(x) = \frac{1}{2(5-3)}$, $3 < x < 5$.

$$\therefore E(X) = 0\times\frac{1}{4} + 1\times\frac{1}{4} + \int_3^5 \frac{1}{2(5-3)}\, dx$$

$$= \frac{1}{4} + \frac{(5-3)}{2(5-3)} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \text{ (Answer)}$$



If distribution function of a mixed random variable given as $F(x) = (1-\lambda) F_1(x) + \lambda F_2(x)$ $0 < \lambda < 1$
where $F_1(x)$ corresponds to discrete distribution & $F_2(x)$ corresponds to continuous distribution
then

$$E(X) = \int_{-\infty}^{\infty} x \, d(F(x))$$

$$= (1-\lambda) \sum_r x_r \cdot P(X=x_r) + \lambda \int_{-\infty}^{\infty} x f_2(x) \, dx$$

Note:- In many cases we have to find expectation of a function of a R.V. instead of the R.V. directly, so in all above working formulae we have to change as below.
if the function is given to be $w(X)$ of the r.v. X. then in all the sums/ integral we have to take weight as $w(x)$ instead of x, so the formulae becomes as below.

① Discrete:- $$E(w(X)) = \sum_{r=-\infty}^{\infty} w(x_r) \cdot P(X=x_r)$$

② Continuous:- $$E(w(X)) = \int_{-\infty}^{\infty} w(x) f(x) \, dx$$

③ Mixed:- $$E(w(X)) = (1-\lambda) \sum_r w(x_r) P(X=x_r) + \lambda \int_{-\infty}^{\infty} w(x) f_2(x) \, dx.$$

Properties of expectation:-

① $E(c) = c$ for any constant c

② $E(X+Y) = E(X) + E(Y)$

③ $E(aX+b) = a E(X) + b$ (By ① & ②)

④ $E(u(X) + v(X)) = E(u(X)) + E(v(X))$ if $E(u(X))$ & $E(v(X))$ exist.

⑤ If $w(x) \geq 0$ $\forall x \in R$ & if $E(w(X))$ exists then $E(w(X)) \geq 0$.

Variance:- The variance of a R.V. X is a measure of how spread out it is. That is less variance means very less spread outside around its expected value while larger variance means lack of consistency towards its expected value and is given by

$Var(X) = E((X - E(X))^2)$ → [Here we wanted to know how X vary around its expected value E(X). You can take any constant value to check how it varies around that constant value]

In general we take variance around its expected value,

Thus $Var(X) = E((X-\mu_X)^2)$ $\mu_X = E(X) = \text{constant}$

$= E(X^2 - 2\mu_X \cdot X + \mu_X^2)$

$= E(X^2) - 2\mu_X E(X) + \mu_X^2$
[By properties of Expectation]

$= E(X^2) - 2\mu_X \cdot \mu_X + \mu_X^2$

$= E(X^2) - \mu_X^2 = E(X^2) - (E(X))^2$

Example:- Let X be a continuous random variable with p.d.f.
$f_X(x) = \begin{cases} \frac{2}{x^2} & \text{for } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$

To find Var(X) we proceed as below.

$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_1^2 x \times \frac{2}{x^2}\,dx = 2\log(2)$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_1^2 x^2 \times \frac{2}{x^2}\,dx = 2$

$\therefore Var(X) = E(X^2) - (E(X))^2 = 2 - (2\log(2))^2$
$= 0.0782.$

Note:- If you can find Expectation, keep faith on you that you can also find variance.
Just work hard on sum / integration.

Also you can use the formulae's as below to find variance when $E(X^2)$ is hard to find directly.

$Var(X) = \sum_r (x_r - m)^2 P(X = x_r)$ (for discrete case) ($m = \mu_X$)
$= \int_{-\infty}^{\infty} (x-m)^2 f(x)\,dx$ (for continuous case), f = p.d.f.

Properties of Variance :-

① If X & Y are independent then $Var(X+Y) = Var(X) + Var(Y)$

② For constants a & b, $Var(aX+b) = a^2 Var(X)$ as $Var(k)=0$ for any const. k.

and any constant random variable is independent of any other random variable.

③ $Var(X) = E(X^2) - (E(X))^2$

④ $Var(X) = E\{X(X-1)\} - m(m-1)$, $m = E(X)$.
when $E(X^2)$ exist. i.e. known to exist.

⑤ $Var(X) = \int_0^\infty 2x\{1 - F(x) + F(-x)\}dx - m^2$, $m = E(X)$

where X is any R.V. & provided $Var(X)$ exist.
where $F(x)$ is its distribution function.

Additional important property of expectation :-

For a continuous random variable X, if $E(X)$ exists (i.e. finite), then

$E(X) = \int_0^\infty \{1 - F(x) - F(-x)\} dx$, F being the distribution function of X.

Significance of the variance of a distribution :

The expression $\sum_r (x_r - m)^2 f_r$ for the variance of X, when X is discrete and the expression $\int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx$ for the variance of X when X is continuous, both give the mean value of $(X-m)^2$ in the long run. So the variance of X gives the mean value of the squares of the deviations of the values of X from the mean m and consequently a low value of Var (X) indicates that there is high concentration of the probability mass near the mean and a high value of the Var (X) indicates that there is low concentration of the probability mass near the mean m. Thus Var (X) gives an inverse measure of concentration of the probability mass near the mean, *i.e.*, if Var (X) is small, then it is highly probable that values of X will be very close to m and if Var (X) is large, then it is highly probable that values of X will deviate much from the mean m. So we can say that Var (X) is a measure of *dispersion* of the distribution.

Again, the expression $\Sigma (x_r - m)^2 f_r$ and $\int_{-\infty} (x-m)^2 f(x)\, dx$ both give the moment of inertia of the distribution of probability mass about the straight line through the centre of mass and perpendicular to the line of distribution in the discrete as also in the continuous case.

Standard Deviation :

The standard deviation of a random variable X, denoted by $\sigma(X)$ or by σ is defined as the non-negative square root of Var(X), so that $\sigma = +\sqrt{\text{Var}(X)}$; *i.e.*, $\sigma^2 = \text{Var}(X)$. (7.4.5)

It is to be noted that the unit of σ is same as that of X, while that of Var (X) is equal to that of the square of X.

Moments :

Let X be a random variable and a be a given real number. The value of $E\{(X-a)^k\}$, if it exists, is called the kth order moment of X (or of the distribution of X) about a where k is a positive integer.

Then $E(X^k)$ is the kth order moment of X about the origin and $E(X^k)$ is denoted by α_k, provided $E(X^k)$ exists. We see that $\alpha_1 = E(X) = m$, the mean of X.

Let the mean $m = E(X)$ exist. Then $E\{(X-m)^k\}$, if it exists, is called the kth order central moment of X and is denoted by μ_k. We observe that $\mu_1 = E(X-m) = E(X) - m = m - m = 0$,

and $\mu_2 = E\{(X-m)^2\} = \text{Var}(X)$.

The moment $E\{(X-a)^k\}$ is also called a raw moment of order k.

The significances of the central moments μ_3, μ_4, etc. will be given at the end of this section.

Relation between raw moments and central moments :

Let X be any random variable having its mean m. If k be any positive integer, by binomial expansion

$$(X-m)^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} X^{k-r} m^r.$$

Then by (7.2.9),

$$E[(X-m)^k]=\sum_{r=0}^{k}(-1)^r\binom{k}{r}E(X^{k-r})\,m^r.$$

$$\therefore\quad \mu_k=\sum_{r=0}^{k}(-1)^r\binom{k}{r}\alpha_{k-r}\,m^r, \qquad (7.4.6)$$

where μ_k and α_k are the kth order central moment and raw moment of order k.

Since $\alpha_0=1$ and $\alpha_1=m$, we get

$$\mu_2=\sum_{r=0}^{2}(-1)^r\binom{2}{r}\alpha_{2-r}\,m^r$$

$$=\alpha_2-2\alpha_1 m+\alpha_0 m^2$$

$$=\alpha_2-2m^2+m^2$$

$$=\alpha_2-m^2,$$

$$\mu_3=\sum_{r=0}^{3}(-1)^r\binom{3}{r}\alpha_{3-r}\,m^r$$

$$=\alpha_3-3\alpha_2 m+3\alpha_1 m^2-\alpha_0 m^3$$

$$=\alpha_3-3\alpha_2 m+2m^3,$$

$$\mu_4=\sum_{r=0}^{4}(-1)^r\binom{4}{r}\alpha_{4-r}\,m^r$$

$$=\alpha_4-4\alpha_3 m+6\alpha_2 m^2-4\alpha_1 m^3+m^4$$

$$=\alpha_4-4\alpha_3 m+6\alpha_2 m^2-3m^4$$

and so on. (7.4.7)

Conversely, $\alpha_2=\mu_2+m^2$,

$$\alpha_3=\mu_3+3\mu_2 m+m^3,$$

$$\alpha_4=\mu_4+4\mu_3 m+6\mu_2 m^2+m^4$$

and so on. (7.4.8)

1. Binomial distribution :

Let X be a binomial (n, p) variate.

Then $m = \text{mean of } X = E(X)$

$$= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=1}^{n} i \frac{n!}{i!\,(n-i)!} p^i (1-p)^{n-i}$$

$$\text{or,}\quad m = np \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!\,(n-i)!} p^{i-1}(1-p)^{n-i}$$

$$= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j}(1-p)^{n-j-1}, j = i-1$$

$$= np\,(p+1-p)^{n-1}$$

$$= np. \tag{7.4.20}$$

Again we have, $E\{X(X-1)\}$

$$= \sum_{i=0}^{n} i(i-1)\binom{n}{i} p^{i}(1-p)^{n-i}$$

$$= n(n-1)p^2 \sum_{i=2}^{n} \binom{n-2}{i-2} p^{i-2}(1-p)^{n-i}$$

$$= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^{j}(1-p)^{n-2-j}, \; j = i-2$$

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

Hence, by (7.4.11)

$$\text{Var}\,(X) = E\{X(X-1)\} - m(m-1)$$

$$= n(n-1)p^2 - np(np-1)$$

$$= np(np-p-np+1)$$

$$= np(1-p). \tag{7.4.21}$$

The corresponding standard deviation $\sigma(X)$ is then given by

$$\sigma(X) = \sqrt{np(1-p)}. \tag{7.4.22}$$

Now we find the raw moments α_2, α_3 and α_4.

$$\alpha_2 = E(X^2) = \sum_{i=0}^{n} i^2 \binom{n}{i} p^{i}(1-p)^{n-i}$$

$$= \sum_{i=0}^{n} \{i(i-1)+i\} \binom{n}{i} p^{i}(1-p)^{n-i}$$

$$= \sum_{i=2}^{n} i(i-1)\frac{n(n-1)}{i(i-1)} \binom{n-2}{i-2} p^{i}(1-p)^{n-i}$$

$$+ \sum_{i=1}^{n} i \binom{n}{i} p^{i}(1-p)^{n-i}$$

$$\therefore\quad \alpha_2 = E(X^2) = n(n-1)p^2\left\{\sum_{i=2}^{n}\binom{n-2}{i-2}p^{i-2}(1-p)^{n-i}\right\}$$

$$+\sum_{i=1}^{n} i\binom{n}{i}p^i(1-p)^{n-i}$$

$$= n(n-1)p^2\sum_{j=0}^{n-2}\binom{n-2}{j}p^j(1-p)^{n-2-j} + E(X)$$

where $j = i-2$

$$= n(n-1)p^2(p+1-p)^{n-2} + np, \quad \text{by (7.4.20)}$$

$$= n(n-1)p^2 + np. \tag{7.4.23}$$

$$\alpha_3 = E(X^3) = \sum_{i=0}^{n} i^3\binom{n}{i}p^i(1-p)^{n-i}$$

$$= \sum_{i=0}^{n}\{i(i-1)(i-2) + 3i(i-1) + i\}\binom{n}{i}p^i(1-p)^{n-i}$$

$$= n(n-1)(n-2)p^3\sum_{i=3}^{n}\binom{n-3}{i-3}p^{i-3}(1-p)^{n-i}$$

$$+3n(n-1)p^2\sum_{i=2}^{n}\binom{n-2}{i-2}p^{i-2}(1-p)^{n-i} + E(X)$$

$$= n(n-1)(n-2)p^3\sum_{j=0}^{n-3}\binom{n-3}{j}p^j(1-p)^{n-3-j}$$

$$+3n(n-1)p^2\sum_{j'=0}^{n-2}\binom{n-2}{j'}p^{j'}(1-p)^{n-2-j'} + np,$$

$j = i-3, j' = i-2.$

$$= n(n-1)(n-2)p^3(p+1-p)^{n-3} + 3n(n-1)(p+1-p)^{n-2}p^2 + np$$

$$= n(n-1)(n-2)p^3 + 3n(n-1)\,p^2 + np. \tag{7.4.24}$$

Similarly writing $i^4 = i(i-1)(i-2)(i-3) + 6i(i-1)(i-2) + 7i(i-1) + i,$

[Put $i^4 = i(i-1)(i-2)(i-3) + Ai(i-1)(i-2) + Bi(i-1) + Ci$ and putting $i = 1, 2, 3$ find A, B, C.]

We get, $\alpha_4 = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np.$

The corresponding central moments μ_2, μ_3, μ_4 then follows from (7.4.7).

$$\mu_2 = \text{Var}(X) = np(1-p). \tag{7.4.25}$$

$$\mu_3 = \alpha_3 - 3\alpha_2 m + 2m^3$$
$$= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np - 3np\{n(n-1)p^2 + np\} + 2n^3p^3$$
$$= np(n^2p^2 - 3np^2 + 2p^2 + 3np - 3p + 1 - 3n^2p^2 + 3np^2 - 3np + 2n^2p^2)$$
$$= np(2p^2 - 3p + 1)$$
$$= np(1-p)(1-2p). \tag{7.4.26}$$

$$\mu_4 = \alpha_4 - 4\alpha_3 m + 6\alpha_2 m^2 - 3m^4$$
$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np - 4np\{n(n-1)(n-2)^2p + 3n(n-1)p^2 + np\} + 6n^2p^2\{n(n-1)p^2 + np\} - 3n^4p^4$$
$$= np(3np^3 - 6np^2 + 3np - 6p^3 + 12p^2 - 7p + 1)$$
$$= np(p-1)\{3np(p-1) - (6p^2 - 6p + 1)\}$$
$$= np(1-p)\{1 + 3p(1-p)(n-2)\}. \tag{7.4.27}$$

Hence, $\beta_2 = \dfrac{\mu_4}{\sigma^4} = \dfrac{np(1-p)\{1+3p(1-p)(n-2)\}}{n^2p^2(1-p)^2}$

$$= \frac{1+3p(1-p)(n-2)}{np(1-p)} = 3 + \frac{1-6p(1-p)}{np(1-p)} \tag{7.4.28}$$

and $\gamma_1 = \dfrac{\mu_3}{\sigma^3} = \dfrac{np(1-p)(1-2p)}{\{np(1-p)\}^{\frac{3}{2}}} = \dfrac{1-2p}{\sqrt{np(1-p)}}$,

$$\beta_1 = \gamma_1^2 = \frac{(1-2p)^2}{np(1-p)} \tag{7.4.29}$$

2. Poisson distribution :

Let X be a poisson μ variate.

Then m = mean of $X = E(X)$

$$= \sum_{i=0}^{\infty} i\, e^{-\mu} \frac{\mu^i}{i!} = \mu e^{-\mu} \sum_{i=1}^{\infty} \frac{\mu^{i-1}}{(i-1)!}$$

$$= \mu e^{-\mu} \sum_{j=0}^{\infty} \frac{\mu^j}{j!}, \qquad j = i-1.$$

$$= \mu e^{-\mu} \cdot e^{\mu} = \mu. \tag{7.4.30}$$

Again, we have $E\{X(X-1)\}$

$$=\sum_{i=0}^{\infty} i(i-1)\, e^{-\mu}\frac{\mu^i}{i\,!}$$

$$=e^{-\mu}\mu^2\sum_{i=2}^{\infty}\frac{\mu^{i-2}}{(i-2)\,!}$$

$$=e^{-\mu}\mu^2\sum_{j=0}^{\infty}\frac{\mu^j}{j\,!},\quad j=i-2$$

$$=e^{-\mu}\mu^2 e^{\mu}=\mu^2.$$

Hence, by (7.4.11)

$$\text{Var}\,(X)=E\{X(X-1)\}-m(m-1)$$

$$=\mu^2-\mu(\mu-1)=\mu. \qquad (7.4.31)$$

The corresponding standard deviation $\sigma(X)$ is then given by

$$\sigma(X)=\sqrt{\mu}. \qquad (7.4.32)$$

We now find the raw moments α_2, α_3 and α_4.

$$\alpha_2=E(X^2)=\sum_{i=0}^{\infty} i^2\,\frac{e^{-\mu}\mu^i}{i\,!}$$

$$=\sum_{i=0}^{\infty}\{i(i-1)+i\}\frac{e^{-\mu}\mu^i}{i\,!}$$

$$=e^{-\mu}\mu^2\sum_{i=2}^{\infty}\frac{\mu^{i-2}}{(i-2)\,!}+E(X)$$

$$=e^{-\mu}\mu^2\sum_{j=0}^{\infty}\frac{\mu^j}{j\,!}+\mu$$

$$=e^{-\mu}\mu^2 e^{\mu}+\mu$$

$$=\mu^2+\mu. \qquad (7.4.33)$$

$$\alpha_3=\sum_{i=0}^{\infty} i^3 \frac{e^{-\mu}\mu^i}{i\,!}=\sum_{i=0}^{\infty}\{i(i-1)(i-2)+3i(i-1)+i\}\frac{e^{-\mu}\mu^i}{i\,!}$$

$$=e^{-\mu}\mu^3\sum_{i=3}^{\infty}\frac{\mu^{i-3}}{(i-3)\,!}+3e^{-\mu}\mu^2\sum_{i=2}^{\infty}\frac{\mu^{i-2}}{(i-2)\,!}+E(X)$$

$$=e^{-\mu}\mu^3\sum_{j=0}^{\infty}\frac{\mu^j}{j\,!}+3e^{-\mu}\mu^2\sum_{j'=0}^{\infty}\frac{\mu^{j'}}{j'\,!}+\mu$$

$$=e^{-\mu}\mu^3e^{\mu}+3e^{-\mu}\mu^2e^{\mu}+\mu$$

$$=\mu^3+3\mu^2+\mu. \qquad (7.4.34)$$

$$\alpha_4=E(X^4)=\sum_{i=0}^{\infty} i^4\frac{e^{-\mu}\mu^i}{i\,!}$$

$$=\sum_{i=0}^{\infty}\{i(i-1)(i-2)(i-3)+6i(i-1)(i-2)+7i(i-1)+i\}\frac{e^{-\mu}\mu^i}{i\,!}$$

$$=e^{-\mu}\mu^4\sum_{i=4}^{\infty}\frac{\mu^{i-4}}{(i-4)\,!}+6e^{-\mu}\cdot\mu^3\sum_{i=3}^{\infty}\frac{\mu^{i-3}}{(i-3)\,!}$$

$$+7e^{-\mu}\mu^2\sum_{i=2}^{\infty}\frac{\mu^{i-2}}{(i-2)\,!}+E(X)$$

$$=e^{-\mu}\mu^4\sum_{j=0}^{\infty}\frac{\mu^j}{j\,!}+6e^{-\mu}\mu^3\sum_{j'=0}^{\infty}\frac{\mu^{j'}}{j'\,!}+7e^{-\mu}\mu^2\sum_{j''=0}^{\infty}\frac{\mu^{j''}}{j''\,!}+\mu$$

$$=e^{-\mu}\mu^4e^{\mu}+6e^{-\mu}\mu^3e^{\mu}+7e^{-\mu}\mu^2e^{\mu}+\mu$$

$$=\mu^4+6\mu^3+7\mu^2+\mu. \qquad (7.4.35)$$

The corresponding central moments μ_2, μ_3, μ_4 then follow from (7.4.7).

$$\mu_2=\text{Var }(X)=\mu, \qquad (7.4.36)$$

$$\mu_3=\alpha_3-3\alpha_2 m+2m^3$$

$$=(\mu^3+3\mu^2+\mu)-3(\mu^2+\mu)\mu+2\mu^3$$

$$=\mu, \qquad (7.4.37)$$

$$\mu_4 = \alpha_4 - 4\alpha_3 m + 6\alpha_2 m^2 - 3m^4$$

$$= (\mu^4 + 6\mu^3 + 7\mu^2 + \mu) - 4\mu(\mu^3 + 3\mu^2 + \mu) + 6\mu^2(\mu^2 + \mu) - 3\mu^4$$

$$= 3\mu^2 + \mu. \quad (7.4.38)$$

Hence, $\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{3\mu^2 + \mu}{\mu^2} = 3 + \frac{1}{\mu}$ (7.4.39)

and $\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu}{\mu^{\frac{3}{2}}} = \frac{1}{\sqrt{\mu}},$ (7.4.40)

$$\beta_1 = \gamma_1^2 = \frac{1}{\mu}.$$

8. **Normal distribution :**

Let X be the random variable having *normal* (m, σ) *distribution.* Then the probability density function $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma > 0, \; -\infty < x < \infty.$$

The mean of normal (m, σ) distribution is then equal to

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} x\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx,$$

provided the integral is absolutely convergent.

Now $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}(x-m)\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$ and $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$

are absolutely convergent and $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx=1.$

Hence, $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}x\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$ is absolutely convergent and its

value is $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}(x-m)\,e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx+m\cdot\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$

$=0+m$

$=m.$

So the mean of normal (m,σ) distribution is equal to m.

Then for any fixed positive integer k,

$$\mu_{2k+1}=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}(x-m)^{2k+1}e^{\frac{-(x-m)^2}{2\sigma^2}}\,dx$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\mathop{Lt}_{\substack{B_2\to\infty\\ B_1\to-\infty}}\int_{B_1}^{B_2}(x-m)^{2k+1}e^{\frac{-(x-m)^2}{2\sigma^2}}\,dx$$

$$=\frac{1}{\sqrt{2\pi}}\mathop{Lt}_{\substack{B_2\to\infty\\ B_1\to-\infty}}\int_{\frac{B_1-m}{\sigma}}^{\frac{B_2-m}{\sigma}}(\sigma z)^{2k+1}e^{\frac{-z^2}{2}}\,dz,$$

where $z=\frac{x-m}{\sigma}$

$$=\frac{\sigma^{2k+1}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^{2k+1}e^{\frac{-z^2}{2}}\,dz=0,$$

since the integrand $z^{2k+1}e^{\frac{-z^2}{2}}$ is an odd function of z and the integral is absolutely convergent.

$\therefore\ \mu_{2k+1}=0,\ k=1,2,3\ldots$ (7.4.58)

Scanned by CamScanner

Again, $\mu_{2k}=\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}(x-m)^{2k} e^{\frac{-(x-m)^2}{2\sigma^2}} dx$

$$=\frac{\sigma^{2k}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{2k} e^{\frac{-z^2}{2}} dz, \text{ as before ;}$$

where $z=\frac{x-m}{\sigma}$

$$=\frac{2\sigma^{2k}}{\sqrt{2\pi}}\int_{0}^{\infty} z^{2k} e^{\frac{-z^2}{2}} dz, \text{ since the integrand } z^{2k} e^{\frac{-z^2}{2}}$$

is an even function of z and the integral is absolutely convergent

$$=\frac{2\sigma^{2k}}{\sqrt{2\pi}} \underset{B\to\infty}{Lt} \int_{0}^{B} z^{2k} e^{\frac{-z^2}{2}} dz \ (B>0)$$

$$=\frac{2\sigma^{2k}}{\sqrt{2\pi}} \underset{B\to\infty}{Lt} \int_{0}^{\frac{B^2}{2}} (2t)^k e^{-t} \frac{dt}{\sqrt{2t}}, \text{ where } \frac{z^2}{2}=t$$

$$=\frac{2^k\sigma^{2k}}{\sqrt{\pi}}\int_{0}^{\infty} t^{k+\frac{1}{2}-1} e^{-t} dt$$

$$=\frac{2^k\sigma^{2k}}{\sqrt{\pi}} \Gamma(k+\tfrac{1}{2}).$$

$$\therefore \ \mu_{2k}=\frac{2^k\sigma^{2k}}{\sqrt{\pi}} \Gamma(k+\tfrac{1}{2})$$

Changing k to $k-1$, we get

$$\mu_{2k-2}=\frac{2^{k-1}\sigma^{2k-2}}{\sqrt{\pi}} \Gamma(k-\tfrac{1}{2}).$$

$$\therefore \ \frac{\mu_{2k}}{\mu_{2k-2}}=2\sigma^2 \cdot \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k-\frac{1}{2})}=2\sigma^2(k-\tfrac{1}{2})=(2k-1)\sigma^2$$

$$[\because \ \Gamma(k+\tfrac{1}{2})=\Gamma(k-\tfrac{1}{2}+1)=(k-\tfrac{1}{2})\,\Gamma(k-\tfrac{1}{2})],$$

i.e., $\mu_{2k}=(2k-1)\sigma^2 \cdot \mu_{2k-2}$

$$=\{(2k-1)\sigma^2\}\{(2k-3)\sigma^2\}\ \mu_{2k-4}$$

$$=\{(2k-1)\sigma^2\}\{(2k-3)\sigma^2\}\{(2k-5)\sigma^2\}\ \mu_{2k-6}$$

$$=\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

$$=\{(2k-1)\sigma^2\}\{(2k-3)\sigma^2\}\ldots 1\sigma^2\mu_0$$

$$=1.3.5\ldots(2k-1)\sigma^{2k} \quad [\because \ \mu_0=1]. \qquad (7.4.59)$$

From above, we then get

$$\text{Var}(X) = \mu_2 = \sigma^2 \quad \text{and} \quad \sigma(X) = \sigma, \tag{7.4.60}$$

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{1.3\sigma^4}{\sigma^4} = 3, \gamma_1 = 0. \tag{7.4.61}$$

Moment generating function :

The *moment generating function* of a random variable X, about a number a, is a real valued function of a real variable t, denoted by M_{X-a} and defined by $M_{X-a} : A \to R$, where $M_{X-a}(t) = E\{e^{t(X-a)}\}$, $t \in A$, provided the expectation exists for $t \in A \subseteq R$, A being the domain of definition of the corresponding moment generating function. We observe that in any case $0 \in A$ for any random variable and so moment generating function is always defined at $t=0$. By the statement "moment generating function exists", we actually mean that moment generating function is defined at least at one point other than zero. Then (*i*) if X be discrete, the moment generating function of X about a is defined by

$$M_{X-a}(t) = \sum_k e^{t(x_k - a)} f_k \tag{7.5.1}$$

provided the series is absolutely convergent and where $P(X = x_k) = f_k$, x_k being a point of the spectrum of X and t belongs to the domain of definition of the moment generating function and (*ii*) if X be continuous, the moment generating function of X about a is defined by

$$M_{X-a}(t) = \int_{-\infty}^{\infty} e^{t(x-a)} f(x)\, dx \tag{7.5.2}$$

provided the integral is absolutely convergent for all t belonging to the domain of definition of the moment generating function and $f(x)$ is the probability density function of X.

For $a=0$, (*i*) if X be discrete, the moment generating function of X, about origin, is defined by $M_X(t) = \sum_k e^{tx_k} f_k$, (7.5.3)

provided the series is absolutely convergent and (*ii*) if X be continuous, the moment generating function of X, about origin, is defined by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, \tag{7.5.4}$$

provided the integral is absolutely convergent, x_k, f_k, $f(x)$ having the usual meanings.

We shall now show that under certain conditions, moments of a random variable X (assuming that the moments exist) about a can be obtained from the corresponding moment generating function $M_{X-a}(t)$.

Case I Let X be discrete. Here

$$M_{X-a}(t)=\sum_k e^{t(x_k-a)} f_k ,$$

provided the series is absolutely convergent. We assume that the above series can be differentiated term by term any number of times with respect to t in the domain of definition A of $M_{X-a}(t)$.

Then $\left[\frac{d^r}{dt^r}\{M_{X-a}(t)\}\right]_{t=0}=\sum_k (x_k-a)^r f_k=\alpha'_r$,

where α'_r is the moment of X about a for $r=1, 2, \ldots$.

Case II. Let X be continuous. Here

$$M_{X-a}(t)=\int_{-\infty}^{\infty} e^{t(x-a)} f(x)\, dx,$$

provided the integral is absolutely convergent. We assume that the integral can be differentiated with respect to t under the integral sign any number of times. Then

$$\left[\frac{d^r\{M_{X-a}(t)\}}{dt^r}\right]_{t=0}=\int_{-\infty}^{\infty} (x-a)^r f(x)dx=\alpha_r$$

for $r=1, 2, 3, \ldots$

Thus we see that under the conditions mentioned above the r-th moment α'_r of a random variable X, discrete or continuous, about a given number a, is given by

$$\alpha'_r=\left[\frac{d^r\{M_{X-a}(t)\}}{dt^r}\right]_{t=0}, \quad r=1, 2, 3, \ldots \tag{7.5.5}$$

If $a=0$, then we write α_r for α_r' and then we get

$$\alpha_r=\left[\frac{d^r M_X(t)}{dt^r}\right]_{t=0}, \quad r=1, 2, 3, \ldots \tag{7.5.6}$$

Now we note that if $M_{X-a}(t)$ can be expanded in a power series in t, then

$$M_{X-a}(t)=\sum_{r=0}^{\infty}\left[\frac{d^r M_{X-a}(t)}{dt^r}\right]_{t=0}\frac{t^r}{r!}, \tag{7.5.7}$$

from which we can state that the r-th order moment of X about a is the coefficient of $\frac{t^r}{r!}$ in the expansion of $M_{X-a}(t)$ as a power series in t.

Some Important Properties :

(i) $M_{X-a}(t)=e^{-at}M_X(t)$, (7.5.8)

provided $M_X(t)$ exists.

(ii) *If c is a constant and Z, U and X are random variables connected by $Z=cX$, $U=c+X$, then*

$$M_Z(t)=M_X(ct),$$
$$M_U(t)=e^{ct}M_X(t), \tag{7.5.9}$$

provided $M_X(t)$ exists.

Proof : (i) $M_{X-a}(t)=E\{e^{t(X-a)}\}$

$=E\{e^{tX}e^{-at}\}=e^{-at}E(e^{tX})$, by (7.2.8)

$=e^{-at}M_X(t)$.

(ii) $M_Z(t)=E\{e^{tZ}\}=E\{e^{(ct)X}\}=M_X(ct)$.

$M_U(t)=E\{e^{tU}\}=E\{e^{t(c+X)}\}=e^{ct}E\{e^{tX}\}$

$=e^{ct}M_X(t)$.

From now, unless otherwise stated, by moment generating function of a random variable X we shall mean moment generating function about zero.

1. Binomial distribution :

Let X be binomial (n, p) variate. Then the moment generating function $M_X(t)$ is given by

$$M_X(t) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k}$$

$$= (1-p+pe^t)^n$$

$$= (pe^t+q)^n, \text{ where } q = 1-p.$$

Thus the moment generating function is given by

$$M_X(t) = (pe^t+q)^n, \text{ for all } t \in R. \tag{7.5.12}$$

We know that in such a case characteristic function can be obtained from $M_X(t)$, replacing t by it.

So the corresponding characteristic function is given by

$$\phi_X(t) = (pe^{it}+q)^n \tag{7.5.13}$$

where $i = \sqrt{-1}$.

2. Poisson distribution :

Let X be a Poisson μ variate. Then the moment generating function $M_X(t)$ is given by

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\mu} \frac{\mu^k}{k!}$$

$$= e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu e^t)^k}{k!}$$

$$= e^{-\mu} e^{\mu e^t} = e^{\mu(e^t-1)}, \text{ for all } t \in R. \tag{7.5.14}$$

Replacing t by it, the corresponding characteristic function is given by

$$\phi_X(t) = e^{\mu(e^{it} - 1)}, \tag{7.6.15}$$

3. Normal distribution :

Let X be a normal (m, σ) variate. Then from (7.5.13) the moment generating function $M_X(t)$ is given by

$$M_X(t) = e^{mt + \frac{1}{2}\sigma^2 t^2}, \text{ for all } t \in R.$$

Then the corresponding characteristic function $\phi_X(t)$ is given by

$$\phi_X(t) = e^{imt - \frac{1}{2}\sigma^2 t^2}. \tag{7.6.16}$$

7.8. Illustrative Examples :

Ex. 1. *A point P is chosen at random on a line segment AB of length 2l. Find the expected values of (i) AP . PB, (ii) | AP − PB |, (iii) max {AP, PB}.*

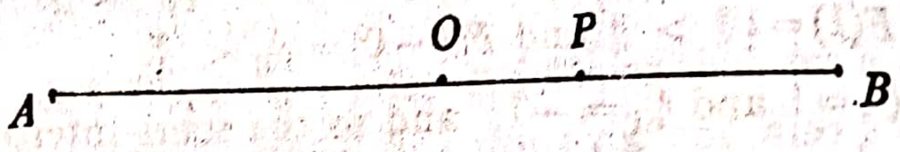


Fig. 7.8.1

Let O be the middle point of AB and X be the random variable denoting the length of OP prefixed with proper sign. Then X has uniform distribution in $(-l, l)$. So if $f(x)$ be the probability density function of X, then

$$f(x)=\frac{1}{2l} \text{ if } -l<x<l$$

$$=0, \text{ elsewhere.}$$

(i) $E\,(AP \cdot PB)=E\,\{(l+X)(l-X)\}=E\,(l^2-X^2)$

$$=\int_{-l}^{l}(l^2-x^2)\cdot\frac{1}{2l}\,dx$$

$$=\frac{1}{2l}\left(2l^3-\frac{2l^3}{3}\right)=\tfrac{2}{3}l^2.$$

(ii) $E\,(\,|\,AP-PB\,|\,)=E\,(\,|\,l+X-l+X\,|\,)=E(\,|\,2X\,|\,)$

$$=\int_{-l}^{l}|\,2x\,|\,\frac{1}{2l}\,dx=\int_{-l}^{l}\frac{|\,x\,|}{l}\,dx$$

$$=\frac{2}{l}\int_{0}^{l}|\,x\,|\,dx=\frac{2}{l}\int_{0}^{l}x\,dx=l.$$

(iii) $E[\max\{AP, PB\}]$

$$=E[\max\{l+X, l-X\}]$$

$$=\int_{-l}^{l} \max\{l+x, l-x\} \frac{1}{2l}\, dx$$

$$=\frac{1}{2l}\int_{-l}^{0} \max\{l+x, l-x\}\, dx+\frac{1}{2l}\int_{0}^{l} \max\{l+x, l-x\}\, dx.$$

Now, $\max\{l+x, l-x\}=l+x$ if $x \geqslant 0$

and $\max\{l+x, l-x\}=l-x$ if $x \leqslant 0$.

So, $E[\max\{l+X, l-X\}]$

$$=\frac{1}{2l}\int_{-l}^{0} (l-x)\, dx+\frac{1}{2l}\int_{0}^{l} (l+x)\, dx$$

$$=\frac{1}{2l}\left(l^2+\frac{l^2}{2}\right)+\frac{1}{2l}\left(l^2+\frac{l^2}{2}\right)=\frac{3l}{2}.$$

Ex. 2. *Find $E(X)$ for the following density function :*

$$f(x)=\frac{4x}{5} \qquad \text{when } 0<x\leqslant 1$$

$$=\tfrac{2}{5}(3-x) \qquad \text{when } 1<x\leqslant 2$$

$$=0, \qquad \text{elsewhere.}$$

[C. H. (Econ.) '91]

$$E(X)=\int_0^2 x\, f(x)\, dx$$

$$=\int_0^1 x\left(\frac{4x}{5}\right) dx+\int_1^2 x\,.\,\tfrac{2}{5}(3-x)\, dx$$

$$=\tfrac{4}{5}\,.\,\tfrac{1}{3}+\tfrac{2}{5}\left(\tfrac{19}{3}-\tfrac{7}{6}\right)=\tfrac{17}{15}.$$

Ex. 3. *If the probability density function of a random variable X is given by $f(x)=C\,e^{-(x^2+2x+3)}$, $-\infty<x<\infty$, find the value of C, the expectation and variance of the distribution.*

[C. H. (Math.) '89]

We have $C\int_{-\infty}^{\infty} e^{-(x^2+2x+3)}\, dx=1$

or, $C\int_{-\infty}^{\infty} e^{-(x+1)^2} e^{-2}\, dx=1$

or, $C\,e^{-2}\int_{-\infty}^{\infty} e^{-z^2}\, dz=1$, $(x+1=z)$

or, $C\,e^{-2}\,\Gamma(\tfrac{1}{2})=1.$

$\therefore\; C=\dfrac{e^2}{\sqrt{\pi}}.$

$$E(X)=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} x\, e^{-(x^2+2x+3)}\, dx$$

$$=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-2} x\, e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x\, e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z-1)\, e^{-z^2}\, dz,\ (x+1=z).$$

Now, $\int_{-\infty}^{\infty} z\, e^{-z^2}\, dz$ is convergent and its value is zero.

$$\therefore\quad E(X)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (-e^{-z^2})\, dz$$

$$=\frac{1}{\sqrt{\pi}}(-\sqrt{\pi})=-1.$$

Again, $E(X^2)=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^2\, e^{-(x^2+2x+3)}\, dx$

$$=\frac{e^2 e^{-2}}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^2 e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z-1)^2 e^{-z^2}\, dz$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z^2-2z+1)e^{-z^2}\, dz.$$

Now, $\int_{-\infty}^{\infty} z^2 e^{-z^2} dz = \underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\int_{B_1}^{B_2} z^2 e^{-z^2}\, dz$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\int_{B_1}^{0} z^2 e^{-z^2} dz+\int_{0}^{B_2} z^2 e^{-z^2}\, dz\right)$$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\int_{B_1^2}^{0} \frac{u\, e^{-u}}{-2\sqrt{u}}\, du+\int_{0}^{B_2^2} \frac{u\, e^{-u} du}{2\sqrt{u}}\right)$$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\frac{1}{2}\int_{0}^{B_1^2} \sqrt{u}\, e^{-u}\, du+\tfrac{1}{2}\int_{0}^{B_2^2} \sqrt{u}\, e^{-u} du\right)$$

$$=\int_{0}^{\infty} u^{\frac{3}{2}-1}\, e^{-u}\, du=\Gamma(\tfrac{3}{2})=\tfrac{1}{2}\sqrt{\pi}.$$

$$\therefore\quad E(X^2)=\frac{1}{\sqrt{\pi}}\left(\tfrac{1}{2}\sqrt{\pi}-0+\sqrt{\pi}\right)=\tfrac{3}{2}.$$

So, $\text{Var}(X)=E(X^2)-\{E(X)\}^2$

$$=\tfrac{3}{2}-(-1)^2=\tfrac{1}{2}.$$

Ex. 4. *If a person gains or loses an amount equal to the number appearing when a balanced die is rolled once according to whether the number is even or odd, how much money can he expect from the game in the long run?* [C. H. (Econ.) '92]

Let X be the random variable denoting the amount of gain or loss as mentioned in the problem. Then the spectrum of X is

$$\{-1, 2, -3, 4, -5, 6\}.$$

Here $P(X=i)=\frac{1}{6}$ for $i=2, 4, 6$ and

$P(X=-i)=\frac{1}{6}$ for $i=1, 3, 5$.

The required expectation is

$$E(X)=(-1+2-3+4-5+6)\,\tfrac{1}{6}=\tfrac{1}{2}.$$

Ex. 5. *If a person gets Rs. $(2x+5)$ where x denotes the number appearing when a balanced die is rolled once, then how much money can be expected in the long run per game?* [C. H. (Econ.) '89]

If X be the random variable denoting the number appearing on the die, then

$$P(X=x)=\tfrac{1}{6}\ \text{for}\ x=1, 2, 3, 4, 5, 6.$$

The required expectation is

$$\sum_{x=1}^{6}(2x+5)\,\tfrac{1}{6}=\tfrac{1}{6}\{2(1+2+3+4+5+6)+30\}=12.$$

So, the person can expect Rs. 12 per game in the long run.

Ex. 6. *Find the expectation of the number of failures preceding the first success in an infinite sequence of Bernoulli trials with probability of success p.*

Let X be the random variable denoting the number of failures preceding the first success. Then the spectrum of X is the enumerable set $\{0, 1, 2, 3, \ldots\}$.

Now, $P(X=i)=(1-p)^i p$ for $i=0, 1, 2, \ldots\ldots$

Then, $$E(X)=\sum_{i=0}^{\infty} i(1-p)^i p = p\sum_{i=1}^{\infty} i(1-p)^i.$$

Now, $$\sum_{i=1}^{\infty} i(1-p)^i$$
$$=\{(1-p)+2(1-p)^2+3(1-p)^3+\ldots\ldots\}$$
$$=(1-p)\{1+2(1-p)+3(1-p)^2+\ldots\ldots\}.$$

We know that the infinite series $1+2x+3x^2+4x^3+\ldots$ is convergent and its sum is $(1-x)^{-2}$ if $|x|<1$.

Now, $0<1-p<1$. Hence, $1+2(1-p)+3(1-p)^2+\ldots$ is absolutely convergent and its sum is $\{1-(1-p)\}^{-2}$.

So, $$E(X)=p(1-p)\{1+2(1-p)+3(1-p)^2+\cdots\}$$
$$=p(1-p)\{1-(1-p)\}^{-2}$$
$$=\frac{p(1-p)}{p^2}=\frac{1-p}{p}.$$

Ex. 7. *A target is fired at and hit 10 times. If the shots are fired independently and if the probability of a hit in each shot is p, find the expectation of shell consumption.*

Let X_i be the random variable denoting the number of shells fired after the $(i-1)$th hit to make the ith hit, for $i=1, 2, \ldots\ldots, 10$.

Then $(X_i=r)$ denotes the event $(r-1)$failures before the rth shell hits the target', where r is a positive integer.

Now, $P(X_i=r)=(1-p)^{r-1}p$ and the spectrum of X_i is the set of all positive integers. Then

$$E(X_i)=\sum_{r=1}^{\infty} r(1-p)^{r-1} p$$
$$=p\{1+2(1-p)+3(1-p)^2+\ldots\ldots\}$$
$$=p\{1-(1-p)\}^{-2}, \text{ since } 0<1-p<1$$
$$=\frac{p}{p^2}=\frac{1}{p}.$$

Thus we see that $E(X_i)=\frac{1}{p}$ and this is true for every i. Hence, the required expectation of shell consumption is

$$E(X_1+X_2+\cdots+X_{10})$$
$$=E(X_1)+E(X_2)+\cdots+E(X_{10}) \quad \text{(see chapter VIII)}$$
$$=10\ E(X_1)=\frac{10}{p}.$$

[It will be proved in the next chapter that if $E(X_1), E(X_2),\ldots, E(X_n)$ exist, then
$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n).]$$

Ex. 8. *If t is a positive real number and the probability mass function of a discrete random variable X is given by*

$$f(x)=e^{-t}\ (1-e^{-t})^{x-1} \text{ for } x=1, 2, 3,\ldots$$
$$=0, \qquad \text{elsewhere,}$$

then find the mean and variance of X.

$$E(X)=\sum_{x=1}^{\infty} e^{-t}\ (1-e^{-t})^{x-1}.\ x$$
$$=e^{-t}\sum_{x=1}^{\infty} x\ (1-e^{-t})^{x-1}$$
$$=e^{-t}\ \{1+2\ (1-e^{-t})+3\ (1-e^{-t})^2+\cdots\}$$
$$=e^{-t}\ \{1-(1-e^{-t})\}^{-2}, \text{ since here } 0<1-e^{-t}<1.$$
$$=e^{-t}\ e^{2t}=e^t.$$

So the required mean is e^t.

Again, $E\{X(X-1)\}$

$$=\sum_{x=1}^{\infty} x\ (x-1)\ e^{-t}\ (1-e^{-t})^{x-1}$$
$$=e^{-t}\sum_{x=2}^{\infty} x(x-1)\ (1-e^{-t})^{x-1}$$
$$=e^{-t}\ \{2.\ 1.\ (1-e^{-t})+3.2\ (1-e^{-t})^2+4.3\ (1-e^{-t})^3+\cdots\}$$
$$=e^{-t}\ (1-e^{-t})\ \{1.2+3.2\ (1-e^{-t})+3.4\ (1-e^{-t})^2+\cdots\},$$

where the infinite series within the second bracket is absolutely convergent for all $t>0$.

Now we know that, for $|x| < 1$,

$$(1-x)^{-1} = 1+x+x^2+x^3+\cdots\cdots \qquad (7.8.1)$$

The right hand side of (7.8.1) being a power series in x, we can differentiate both sides of (7.8.1) any number of times for $|x| < 1$.

Differentiating twice with respect to x both sides of (7.8.1) we get

$$\frac{2}{(1-x)^3} = 2.1+3.2.\ x+4.3.\ x^2+\cdots$$

Now, $0 < 1-e^{-t} < 1$, since we have $t > 0$.

So, $1.2+2.3\ (1-e^{-t})+3.4.\ (1-e^{-t})^2+\cdots\cdots$

$$= \frac{2}{\{1-(1-e^{-t})\}^3} = 2e^{3t}.$$

Hence, we get $E[X(X-1)] = e^{-t}\ (1-e^{-t})\ 2e^{3t}$

$$= 2e^{2t}\ (1-e^{-t}).$$

So the required variance is $E\ [X(X-1)] - m(m-1)$

$$= 2e^{2t}\ (1-e^{-t}) - e^t\ (e^t-1) \qquad [\because\ m = E(X) = e^t]$$

$$= e^{2t} - e^t.$$

Ex. 9. *A special die with $n+1$ faces is marked in its faces the numbers $0, \frac{1}{n}, \frac{2}{n}, \ldots\ldots, \frac{n-1}{n}, \frac{n}{n}$. The die is unbiased. Let X be the random variable denoting the number on the uppermost face. Find (a) $E(X)$, (b) $Var(X)$ and (c) coefficient of skewness of the distribution of X.*

Here $P\left(X = \frac{i}{n}\right) = \frac{1}{n+1}$ for $i = 0, 1, 2, \cdots\cdots n$.

$$E(X) = \sum_{i=0}^{n} \frac{i}{n} \cdot \frac{1}{n+1} = \frac{1}{n(n+1)}\ (1+2+\cdots+n) = \tfrac{1}{2}.$$

$$\text{Var}\ (X) = E(X^2) - \{E(X)\}^2$$

$$= \frac{1}{n+1} \sum_{i=0}^{n} \frac{i^2}{n^2} - \tfrac{1}{4}.$$

$$= \frac{1}{n^2(n+1)} \cdot \frac{n(n+1)(2n+1)}{6} - \tfrac{1}{4}$$

$$= \frac{2n+1}{6n} - \tfrac{1}{4} = \frac{n+2}{12n}.$$

(c) Coefficient of skewness is given by

$$\gamma_1 = \frac{\mu_3}{\sigma^3}.$$

Now, $\mu_3 = E\{(X-\frac{1}{2})^3\}$

$$= E(X^3) - \frac{3}{2} E(X^2) + \frac{3}{4} E(X) - \frac{1}{8}$$

$$= \frac{1}{n+1}\sum_{i=0}^{n}\left(\frac{i}{n}\right)^3 - \frac{3}{2}\,\frac{2n+1}{6n} + \frac{3}{8} - \frac{1}{8}$$

$$= \frac{1}{n^3(n+1)}\left\{\frac{n(n+1)}{2}\right\}^2 - \frac{2n+1}{4n} + \frac{1}{4}$$

$$= \frac{n+1}{4n} - \frac{2n+1}{4n} + \frac{1}{4}$$

$$= \frac{n+1-2n-1+n}{4n} = 0.$$

Hence, $\gamma_1 = 0$.

Ex. 10. *Find the mean, variance and the coefficient of skewness of the continuous distribution with probability density function given by*

$$f(x) = 1 - |1-x|,\ 0 < x < 2$$
$$= 0 \qquad ,\ \textit{elsewhere}.$$

If X be the corresponding random variable, then the mean is

$$E(X) = \int_0^2 (1 - |1-x|)\, x\, dx$$

$$= \int_0^1 (1-|1-x|)\, x\, dx + \int_1^2 (1-|1-x|)\, x\, dx$$

$$= \int_0^1 \{1-(1-x)\}\, x\, dx + \int_1^2 \{1-(x-1)\}\, x\, dx$$

$$= \int_0^1 x^2\, dx + \int_1^2 (2x - x^2)\, dx$$

$$= \frac{1}{3} + \frac{2}{3} = 1.$$

Now $E(X^2)=\int_0^2 x^2 (1-|1-x|)\,dx$

$$=\int_0^1 x^2 (1-|1-x|)\,dx+\int_1^2 x^2 (1-|1-x|)\,dx$$

$$=\int_0^1 x^2(1-1+x)\,dx+\int_1^2 x^2(1-x+1)\,dx$$

$$=\int_0^1 x^3\,dx+\int_1^2 (2x^2-x^3)\,dx$$

$$=\tfrac{1}{4}+(\tfrac{16}{3}-\tfrac{16}{4})-(\tfrac{2}{3}-\tfrac{1}{4})$$

$$=\tfrac{1}{4}+\tfrac{16}{12}-\tfrac{5}{12}=\tfrac{7}{6}.$$

So, var $(X)=E(X^2)-\{E(X)\}^2=\frac{7}{6}-1=\frac{1}{6}$.

$$\mu_3=E\{(X-1)^3\}$$

$$=E(X^3)-3E(X^2)+3E(X)-1$$

$$=E(X^3)-3.\tfrac{7}{6}+3-1$$

$$=E(X^3)-\tfrac{7}{2}+2$$

$$=\int_0^1 x^3(1-1+x)\,dx+\int_1^2 x^3(1-x+1)\,dx-\tfrac{3}{2}$$

$$=\int_0^1 x^4\,dx+\int_1^2 (2x^3-x^4)\,dx-\tfrac{3}{2}$$

$$=\tfrac{1}{5}+\tfrac{8}{5}-\tfrac{3}{10}-\tfrac{3}{2}=0.$$

So, coefficient of skewness is

$$\gamma_1=\frac{\mu_3}{\sigma^3}=0.$$