Probability theory.

Random Experiment – It may be defined as an experiment which when repeated under essentially identical conditions does not give unique results but may result in any one of the several possible outcomes. These outcomes are known as events or cases.

Exhaustive Events – The total no of possible outcomes in a random experiment are known as Exhaustive events.

Favorable Events – The events which cause the happening of a particular event A are called the favorable events to the event A.

Mutually Exclusive Events – The events, where occurence of one rules out the occurance of the other, are called mutually exclusive events.

Equally Likely events – The events are said to be equally likely if none of them is expected to occur in preference to other

Mathematical or Classical definition of Probability –

If there are n exhaustive, mutually exclusive & equally likely events, out of which m are favourable to the happening of of an event A, then the probability of happening of A, denoted by P(A) is defined as

$$P(A) = \frac{\text{Favourable no. of cases}}{\text{Exhaustive no. of cases}} = \frac{m}{n}$$

The probability that A will not happen will be

$$P(\overline{A}) = \frac{n-m}{n} = 1 - P(A)$$

Since $0 \le m \le n$ so $0 \le \frac{m}{n} \le 1$

If $P(A) = 0$, A is called impossible event

If $P(A) = 1$, A is called certain or sure event.

P(A) is called probability of success & usually denoted by p & $1 - P(A)$ i.e $P(\overline{A})$ is called probability of failure & is usually denoted by q.

e.g : In a single throw of two dice, find the probability of getting a total of 10.

solu : Total events $= 6 \times 6 = 36$

favourable events = (4, 6), (6, 4), (5, 5)

$\therefore \; P(A) = \frac{3}{36} = \frac{1}{12}$

e.g : from a pack of 52 cards, two cards are drawn at random. find the chance that one is King & other a queen.

solu : Total no. of cases $= {}^{52}C_2$.

favarable cases $= {}^4C_1 \times {}^4C_1$

$\therefore \; P(A) = \frac{{}^4C_1 \times {}^4C_1}{{}^{52}C_2} = \frac{8}{663}$

eg : A bag contain 7 white, 6 red & 5 black balls. Two balls are drawn at random. Find the probability that they both are white

Solu : Total cases $= {}^{18}C_2$ (Balls are 18 in no.)

favourable cases $= {}^7C_2$

$\therefore \; P(A) = \frac{{}^7C_2}{{}^{18}C_2} = \frac{7}{51}$

Sample space – The set of all possible outcomes of a random experiment is called a sample space of the experiment. The sample space is denoted by S.

A sample space is discrete if it ~~is~~ consist of a finite or countable infinite set of outcomes. A sample space is continuous if it contains an interval of real numbers. e.g: A sample space consisting of all real no. is a continuous sample space, whereas a sample space consisting of only (yes, no) is a discrete sample space.

Event – An event is a subset of the sample space of a random experiment.

The union of two events is the event that consists of all outcomes that are contained in either of the two events. It is denoted by $E_1 \cup E_2$.

The intersection of two events is the event that consists of all outcomes that are contained in both of the two events. we denote the intersection as $E_1 \cap E_2$.

The complement of an event in a sample space is the set of outcomes in the sample space that are not in the event. we denote the complement of E as E' or E^c.

e.g: $S = \{yy, nn, yn, ny\}$ is a sample space

$E_1 = \{yy, yn, ny\}$ is an event

Let $E_2 = \{yy\}$ & $E_3 = \{yy, yn\}$

Then $E_2 \cup E_3 = \{yy, yn\}$

$E_2 \cap E_3 = \{yy\}$

$\bar{E}_2 = E_2^c = \{nn, yn, ny\}$

for mutually exclusive events E_1 & E_2

$$E_1 \cap E_2 = \phi \quad \text{(null set)}$$

We also have $(E^c)^c = E$

By set theory we have

$$(A \cup B)' = A' \cap B'$$

$$(A \cap B)' = A' \cup B'$$

& $A \cap B = B \cap A$ & $A \cup B = B \cup A$.

Counting Techniques –

① Multiplication rule –

Assume an operation can be described as a sequence of K steps and

the no. of ways of completing step 1 is n_1, and

the no. of ways of completing step 2 is n_2, for each way of completing step 1, and

the no. of ways of completing step 3 is n_3, for each way of completing step, and

so forth.

The total number of ways of completing the operation is

$$n_1 \times n_2 \times \ldots \times n_K.$$

② Permutations – Consider a set of elements such as $S = (a, b, c)$. A permutation of the elements is an ordered sequence of the elements. e.g, abc, cab, acb, bac, bca & cba are all of the permutations of the elements of S.

The number of permutations of n different elements is $n!$, where $n! = n \times (n-1) \times (n-2) \times \ldots \times 1$

Note: The number of permutations of subsets of r elements selected from a set of n different elements is

$$P_r^n = \frac{n!}{(n-r)!}$$

③ Combinations – It is similar to permutation, but here order is not important. So, the no. of combinations, subsets of size r that can be selected from a set of n elements is, denoted as ^nC_r &

$$^nC_r = \frac{n!}{r!\,(n-r)!}$$

Interpretations & Axioms of Probability–

The probability of an outcome can be interpreted as our subjective probability, or degree of belief, that the outcome will occur. Different individuals with no doubt assign different probabilities to the same outcomes.

In terms of sample space, the probability is defined as

for a discrete sample space, the probability of an event E, denoted as P(E) equals the sum of the probabilities of the outcomes in E.

Whenever the sample space consists of N possible outcom that are equally likely, the probability of each outcome is 1/N.

Axioms of Probability – Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties

If S is a sample space & E is any event in the random experiment,

1) $P(S)=1$

2) $0 \le P(E) \le 1$

3) For two events E_1 & E_2 with $E_1 \cap E_2 = \phi$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

The axioms implies that

$$P(\phi) = 0$$

& for any event E, $P(E') = 1 - P(E)$

If the event E_1 is contained in the event E_2

$$P(E_1) \le P(E_2)$$

Union & addition of Event –

If A & B are any two events (subsets of sample spaces) and are not disjoint, then

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

Since for two mutually exclusive events A & B

$$A \cap B = \phi \Rightarrow P(A \cap B) = 0 \quad \therefore$$

$$P(A \cup B) = P(A) + P(B)$$

Three or more events –

$$\begin{aligned} P(A \cup B \cup C) &= P[(A \cup B) \cup C] \\ &= P(A \cup B) + P(C) - P[(A \cup B) \cap C] \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P[(A \cap C) \cup (B \cap C)] \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

for a collection of events $E_1, E_2 \ldots, E_n$ which are mutually exclusive

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$$

Conditional Probability –

The conditional probability of an event B given an event A, denoted by $P(B/A)$ is

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \quad \text{for } P(A) > 0$$

Multiplication & Total Probability rules –

from conditional probability, we can write

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

For any events A & B

$$P(B) = P(B \cap A) + P(B \cap A')$$

$$= P(B/A)\,P(A) + P(B/A')\,P(A')$$

Total probability rule for multiple events –

Assume $E_1, E_2 \ldots E_k$ are k mutually exclusive & exhaustive sets. Then

$$P(B) = P(B \cap E_1) + P(B \cap E_2) + \ldots + P(B \cap E_k)$$

$$= P(B/E_1)\,P(E_1) + P(B/E_2)\,P(E_2) + \ldots + P(B/E_k)\,P(E_k)$$

Independent Events – Two events are independent if any one of the following equivalent statements is true:

(1) $P(A/B) = P(A)$

(2) $P(B/A) = P(B)$

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

Note: The events $E_1, E_2 \ldots E_n$ are independent if & only if for any subset of these events $E_{i_1}, E_{i_2}, \ldots E_{i_K}$

$$P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_K}) = P(E_{i_1}) \times P(E_{i_2}) \times \ldots \times P(E_{i_K})$$

Result: If p_1, p_2 be two probabilities of happening of two independent events, then

1) the probability that the first event happen & the second fails is $p_1(1-p_2)$

2) the probability that the first & second both events fail to happen is $(1-p_1)(1-p_2)$

3) the prob. that at least one of the event happens is $1-(1-p_1)(1-p_2)$. This is called cumulative probability

In general, if $p_1, p_2 \ldots p_n$ be the chances of happening of n independent events, then their cumulative probability i.e the chance that at least one of the event will happen is $1-(1-p_1)(1-p_2)\ldots(1-p_n)$.

eg: Find the chance of obtaining at least one 6 in a throw of four dice.

solu: Prob of getting $6 = 1/6$

" " Not getting $6 = 5/6$

Prob of getting at least one 6 in a throw of four dice is $= 1-\left(\frac{5}{6}\right)^4$

example – The probability that a teacher will give an unannounced test during any class meeting is 1/5. If a student is absent twice, what is the prob. that he will miss at least one test?

Solu: The student will not miss any test if the teacher does not give any test on two days when he is absent.

The prob. that the teacher will not give any test on the two days $= \left(1-\frac{1}{5}\right)\left(1-\frac{1}{5}\right) = \frac{16}{25}$

The prob that the student will miss at least one test $= 1-\frac{16}{25} = \frac{9}{25}$.

xample – Two persons A and B toss an unbiased coin alternately on the understanding that the first who gets the head wins. If A starts the game, find their respective chances of winning.

Solu: Prob. of getting a head = 1/2

Then A can win in 1st, 3rd, 5th, ... throws.

$\therefore$ Chances that A wins $= \frac{1}{2} + \left(\frac{1}{2}\right)^2 \frac{1}{2} + \left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) + \dots$

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

$\therefore$ Chance of B's winning $= 1-\frac{2}{3} = 1/3$.

example: A box A contains 2 white & 4 black balls. Another box B contains 5 white & 7 black balls. A ball is transferred from box A to box B. Then a ball is drawn from box B. what is the prob. that it is white?

Solu: The prob. of drawing a white ball from box B will depend on whether the transferred ball is black or white. If black ball is transferred its prob is 4/6 Then Box B has 5 white & 8 black balls. Now the prob of drawing white ball is $\frac{5}{13}$. ∴ prob of drawing white ball in this case is $\frac{4}{6} \times \frac{5}{13} = \frac{10}{39}$.

If white ball is transferred its prob is 2/6 & then B has 6 white & 7 black balls. Now the prob of drawing white ball is $\frac{6}{13}$. ∴ Prob of drawing white ball in this case is $\frac{2}{6} \times \frac{6}{13} = \frac{2}{13}$

Total prob $= \frac{10}{39} + \frac{2}{13} = \frac{16}{39}$.

Example: A can hit a target 3 times in 5 shots, B 2 times in 5 shots, C 3 times in 4 shots. Find the probability that (i) two shot hits (ii) atleast two shot hits, when they fire together.

solu: Prob. of A hitting target = 3/5
" " B " " = 2/5
" " C " " = 3/4.

(i) A & B hit the target, C fails $= \frac{3}{5} \times \frac{2}{5} \times \frac{1}{4} = \frac{6}{100} = P(1)$
B & C " " " , A fails $= \frac{2}{5} \times \frac{2}{5} \times \frac{3}{4} = \frac{12}{100} = P(2)$
A & C " " " , B fails $= \frac{3}{5} \times \frac{3}{5} \times \frac{3}{4} = \frac{27}{100} = P(3)$

∴ two shot hits if one of these occurs
$= P(1) + P(2) + P(3)$

$= \frac{6}{100} + \frac{12}{100} + \frac{27}{100} = \frac{45}{100}$

(ii) At least two shot hits if in addition to above events, One more event occur which is all three hit the target.

$$P(\text{All hit the target}) = P(4) = \frac{3}{5} \times \frac{2}{5} \times \frac{3}{4} = \frac{18}{100}$$

$$\therefore P(\text{at least two shot hits}) = \frac{45}{100} + \frac{18}{100} = \frac{63}{100}$$

Huyghen's Problem – A & B throw with a pair of dice. A wins if he throws 6 before B throws 7 & B wins if he throws 7 before A throws 6. If A begins, show that his chance of winning is 30/61.

Solu: Prob that A gets 6 $= \frac{5}{36}$ (1,5) (5,1) (2,4) (4,2) (3,3)

Prob that B get 7 $= \frac{6}{36}$ (1,6) (6,1) (2,5) (5,2) (3,4) (4,3) (

$P(\bar{A}) = \frac{31}{36}$ $P(\bar{B}) = \frac{30}{36}$

A wins in following ways :

$A, \bar{A}\bar{B}A, \bar{A}\bar{B}\bar{A}\bar{B}A, \ldots$ & so on

$$P(A \text{ wins}) = \frac{5}{36} + \frac{31}{36}\,\frac{30}{36}\,\frac{5}{36} + \left(\frac{31}{36}\right)^2\left(\frac{30}{36}\right)^2\frac{5}{36} + \ldots$$

$$= \frac{5}{36}\;\frac{1}{1 - \frac{31}{36}\cdot\frac{30}{36}} = \frac{30}{61}$$

Baye's theorem – If an event E can occur in combination with one of the mutually exclusive events $E_1, E_2 \ldots E_n$, then

$$P(E_i/E) = \frac{P(E_i)\,P(E/E_i)}{\sum_{i=1}^{n} P(E_i)\,P(E/E_i) = P(E)} \qquad i = 1, 2 \ldots n$$

The probabilities $P(E_i)$; $i = 1, 2 \ldots n$ are called apriori probabilities because these exist before we get any information from the experiment

The probabilities $P(E/E_i)$ are called posteriori probabilities because these are found after the experiment results are known.

example – In a bolt factory machines A, B, C manufactures respectively 25, 35 and 40 percent of the total. Out of their output 5, 4 & 2 percent are defective bolts. A bolt is drawn from the produce & is found defective. What are the probabilities that it was manufactured by A, B & C?

Solu: Let E be the event that bolt is defective. & E_1, E_2 & E_3 as events that bolt is produced by A, B & C respectively. Then

$$P(E_1) = 25\% = 0.25$$
$$P(E_2) = 0.35$$
$$P(E_3) = 0.40$$

Now $P(E/E_1)$ = Prob that the bolt produced by A is defective $= 0.05$

$$P(E/E_2) = 0.04$$
$$P(E/E_3) = 0.02$$

Now prob. that the bolt drawn & found defectives was manufactored by A =

$$P(E_1/E) = \frac{P(E_1)\,P(E/E_1)}{\sum_{i=1}^{3} P(E_i)P(E/E_i)}$$

$$= \frac{0.25 \times 0.05}{0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02}$$

$$= 25/69$$

IIIy $P(E_2/E) = 28/69$

$P(E_3/E) = 16/69$

example– Urn A contains 2 white & 2 black balls & Urn B contains 3 white & 2 black balls. One ball is transferred from A to B & then one ball is drawn out of B. Find the chance that this ball is white. If this ball turns out to be white, find the prob. that the transferred ball was white.

Solu: Let E_1 & E_2 are events that the transferred ball from A to B is white & black respectively.
Let W be the event that white ball is drawn from B.

$P(E_1) = \frac{2}{4}$ $\quad P(E_2) = \frac{2}{4}$

$P(W/E_1) = \frac{4}{6}$ $\quad P(W/E_2) = \frac{3}{6}$

$$P(W) = P(E_1)P\left(\frac{W}{E_1}\right) + P(E_2)P\left(\frac{W}{E_2}\right)$$
$$= \frac{2}{4}\cdot\frac{4}{6} + \frac{2}{4}\cdot\frac{3}{6} = \frac{7}{12}.$$

Now
$$P(E_1/W) = \frac{P(E_1)P(W/E_1)}{P(W)} = \frac{\frac{2}{4}\cdot\frac{4}{6}}{\frac{7}{12}} = \frac{4}{7}$$

example :– suppose 5 men out of 100 & 25 women out of 10,000 are colour blind. A colour blind person is chosen at random. Find the probability of his being male. (Assume male & female are equal numbers)

Solu: E_1 = Event that person is male $\quad$ E = person is colour blind
E_2 = " " " is female

$P(E_1) = P(E_2) = 1/2$ $\quad$ Now $P(E|E_1) = 0.05$

$P(E/E_2) = 0.0025$ $\quad$ Now

$$P(E_1/E) = \frac{P(E_1)P(E/E_1)}{\sum_{i=1}^{2} P(E_i)P(E|E_i)} = \frac{\frac{1}{2}\times 0.05}{\frac{1}{2}\times 0.05 + \frac{1}{2}\times 0.0025}$$

Random Variate –

Suppose a variate x takes n values $x_1, x_2, \ldots x_n$ with their probabilities $p_1, p_2 \ldots p_n$ respectively. Since the values of the variate x depends on chance, it is called discrete random variate or stochastic variate. e.g. if we throw a die, then the variate x takes the values 1, 2, 3, 4, 5, 6 each with probability 1/6 and so x is a discrete random variate.

On the other hand if a variate x takes infinitely many values with respective probabilities then it is called continous variate.

If in a random experiment, the event corresponding to a number c occurs, then the corresponding random variable x is said to assume value c & the probability of the event is denoted by $P(x=c)$. Similarly, the probability of the event $x \leq c$ is written as $P(x \leq c)$.

Discrete Probability Distibution –

Suppose a discrete variate X is the outcome of some experiment. If the probability that X takes the values x_i, is p_i, then

$$P(X=x_i) = p_i \quad \text{or} \quad p(x_i) \text{ for } i = 1, 2 \ldots$$

where $p(x_i) \geq 0 \ \forall i$

and $\sum_i p(x_i) = 1$.

The set of values x_i with their probabilites p_i constitute a discrete probability distibution of the discrete variate X.

eg: The discrete prob. dist. for X, the sum of numbers on dices (pair) after they tossed is

$X=x_i$	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_i)$	$1/36$	$2/36$	$3/36$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Distribution function – The D.F $F(x)$ of the discrete variate X is defined by

$$F(x) = P(X \leq x) = \sum_{i=1}^{x} p(x_i) \quad ; x \text{ is any integer.}$$

The distribution function is also called cumulative distribution function.

example – A random variable X has the following probability function

$X=x$:	0	1	2	3	4	5	6	7
$P(X=x)$:	0	K	2K	2K	3K	K^2	$2K^2$	$7K^2+K$

(i) Find K (ii) Evaluate $P(X<6)$, $P(X \geq 6)$ & $P(0<X<5)$

(iii) If $P(X \leq a) > \frac{1}{2}$, find minimum value of a.

(iv) Find Distribution function of X.

Solu : (i) Since $\sum_i P(X=x_i) = 1$ $\therefore$

$$0+K+2K+2K+3K+K^2+2K^2+7K^2+K = 1$$

$$K = 1/10$$

(ii) $P(X<6) = P(X=0)+P(X=1)+\ldots+P(X=5)$

$$= 0+\frac{1}{10}+\frac{2}{10}+\frac{2}{10}+\frac{3}{10}+\frac{1}{100} = \frac{81}{100}$$

$P(X \geq 6) = 1-P(X<6)$

$$= 1-\frac{81}{100} = \frac{19}{100}$$

$$P(0<X<5) = P(X=1) + P(X=2) + \dots + P(X=3)$$
$$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} = \frac{8}{10}$$

(iii) If $P(X \le a) > \frac{1}{2}$ then $P(X \le 3) = \frac{1}{10} + \frac{2}{10} + \frac{2}{10} = \frac{1}{2}$

$\therefore$ $P(X \le 4) > \frac{1}{2}$ $\therefore$ $a = 4$.

(iv)

X	$F(x) = P(X \le x) = \sum_{i=1}^{x} P(x_i)$
0	0
1	$K = 1/10$
2	$K + 2K = 3K = 3/10$
3	$K + 2K + 2K = 5/10$
4	$K + 2K + 2K + 3K = 8/10$
5	$K + 2K + 2K + 3K + K^2 = 8K + K^2 = 81/100$
6	$8K + 3K^2 = 83/100$
7	$9K + 10K^2 = 1$.

The distribution fn or cumulative distribution fn $F(x)$ satisfies the following properties

1) $F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$

2) $0 \le F(x) \le 1$

3) If $x \le y$, then $F(x) \le F(y)$

Mean & Variance of a random variable –

The mean or expected value of the discrete random variable X, denoted as μ or E(X), is

$$\mu = E(X) = \sum_x x f(x)$$

The variance of X, denoted by σ^2 or V(X), is

$$\sigma^2 = V(X) = E(X-\mu)^2$$

$$= \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

The standard deviation of X is $\sigma = \sqrt{\sigma^2}$

Proof: $V(X) = E(X-\mu)^2$

$$= \sum (x^2 + \mu^2 - 2\mu x) f(x)$$

$$= \sum x^2 f(x) + \mu^2 \sum f(x) - 2\mu \sum x f(x)$$

$$= \sum x^2 f(x) + \mu^2 - 2\mu.\mu = \sum x^2 f(x) - \mu^2$$

(Since $\sum f(x) = 1$)

Expected value of a function of a discrete random variable –

If X is a discrete random variable with p.m.f f(x)

$$E[h(X)] = \sum_x h(x) f(x)$$

$$\therefore Var(X) = \sum x^2 f(x) - \mu^2 = E(X^2) - [E(X)]^2$$

Discrete Uniform distribution – A random variable X has a discrete uniform distribution if each of the n values in its range, say, $x_1, x_2 \ldots x_n$, has equal probability. Then

$$f(x_i) = 1/n$$

Suppose the range of the discrete random variable X is the consecutive integers $a, a+1, a+2, \dots b$ for $a \le b$. The range of X contains $b-a+1$ values each with probability $\frac{1}{(b-a+1)}$. Now

$$\mu = \sum_{k=a}^{b} k\left(\frac{1}{b-a+1}\right) = \frac{1}{b-a+1}\sum_{k=a}^{b} k$$

$$= \frac{1}{b-a+1} \times \frac{b(b+1)-a(a-1)}{2}$$

$$\text{Mean } \mu = \frac{b^2+b-a^2+a}{2(b-a+1)} = \frac{(a+b)(b-a+1)}{2(b-a+1)} = \frac{a+b}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b-a+1)^2-1}{12}$$

Ex: Let the random variable X denote the number of the 48 voice lines that are in use at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Then

$$\mu = \frac{48+0}{2} = 24 \quad \&$$

$$\sigma^2 = \frac{(48-0+1)^2-1}{12} = \frac{49^2-1}{12} = \frac{2400}{12} = 200$$

Bernoulli's distribution —

A random variate X which takes values 0 and 1 with probabilities q & p (for failure & success resp.) i.e

$$P(X=1) = p$$

$$P(X=0) = q = 1-p$$

is called Bernoulli's variate. The probability distribution of variate X is called Bernoulli's distribution.

Binomial Distribution— If we perform a series of independent trials such that for each trial p is the probability of a success and q that of a failure, then the prob. of r successes in a series of n trials is given by $^nC_r p^r q^{n-r}$ where r takes integral values from 0 to n. The prob. of 0, 1, 2, ... r, ... n succeses are therefore given by

$^nC_0 p^0 q^n$, $^nC_1 p q^{n-1}$, $^nC_2 p^2 q^{n-2}$, ... $^nC_r p^r q^{n-r}$, ... $^nC_n p^n q^0$

The probabilities of the number of successes so obtained is called the binomial prob. distribution because the probabilities are the successive terms in the expansion of binomial $(q+p)^n$.

∴ The sum of probabilities =

$$q^n + {}^nC_1 p q^{n-1} + \dots + p^n = (q+p)^n = 1.$$

The binomial distribution contains two independent parameters n & p (or q).

Suppose n independent trials constitute one experiment. If this experiment is repeated N times, then the probability of x successes is

$$P(x) = N\,{}^nC_x p^x q^{n-x} \quad x = 0, 1, 2, \dots n$$

These prob. are given in the Binomial expansion of

$$N(q+p)^n$$

This is called the Binomial frequency distribution.

eg: An experiment succeedes twice as often as it fails. Find the chance that in the next six trials there will be at least 4 successes.

Solu: The prob. of getting success $= p = 2/3$

" " " " failure $= q = 1/3$.

Prob of getting x successes $= {}^6C_x \, p^x q^{6-x} = P(x)$

" " " at least 4 success $= P(4)+P(5)+P(6)$

$$= {}^6C_4 \, p^4 q^2 + {}^6C_5 \, p^5 q + {}^6C_6 \, p^6$$

$$= {}^6C_4 \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^2 + {}^6C_5 \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right) + {}^6C_6 \left(\frac{2}{3}\right)^6$$

$$= \frac{496}{729}$$

eg: If on an average 1 vessel in every 10 is wrecked, find the probability that out of 5 vessels expected to arrive, at least 4 will arrive safely.

Solu: prob. of vessel arrive safely, $p = \frac{9}{10}$

" " " wrecked, $q = \frac{1}{10}$

$\therefore$ Prob that x vessel out of 5 arrive safely =

$$P(x) = {}^5C_x \, p^x q^{5-x}$$

$\therefore$ Prob. that at least 4 arrive safely $= P(4)+P(5)$

$$= {}^5C_4 \left(\frac{9}{10}\right)^4 \left(\frac{1}{10}\right) + {}^5C_5 \left(\frac{9}{10}\right)^5$$

$$= 0.91854.$$

e.g: Six dices are thrown 729 times. How many times do you expect at least three dice to show a 5 or 6?

Solu: prob of occurance of 5 or 6 on a die $= \frac{2}{6} = \frac{1}{3} = p$

" " non " " " $q = \frac{2}{3}$.

$n = 6$ $\therefore$ prob of x success is given by =

$$P(x) = {}^6C_x \, p^x q^{6-x}.$$

Prob that at least three dice show 5 or 6 (in one throw)

$$= P(3)+P(4)+P(5)+P(6)$$

$$= {}^6C_3\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^3 + {}^6C_4\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2 + {}^6C_5\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right) + {}^6C_6\left(\frac{1}{3}\right)^6$$

$$= \frac{233}{729}$$

Hence the number of times at least 3 successes occur

$$= \frac{233}{729} \times 729 = 233.$$

$$\sum_{x=0}^{n} x\, {}^nC_x p^x q^{n-x}$$

$$np\sum_{x=0}^{n} x \frac{(n-1)!}{(n-x)!\,x!} p^{x-1} q^{n-x}$$

$$= np\left[0 + 1.\frac{(n-1)!}{(n-1)!\,1!} p^0 q^{n-1} \ldots\right.$$

$$np\sum_{x=1}^{n-1} {}^{n-1}C_{x-1} p^x q^{n-x}$$

Moments of B.D. —

The expectation/mean of B.D is

$$\mu = E(x) = \sum_{x=0}^{n} x\, {}^nC_x\, p^x q^{n-x}$$

$$= {}^nC_1\, p\, q^{n-1} + 2\,{}^nC_2\, p^2 q^{n-2} + \ldots + n\,{}^nC_n\, p^n \quad np[$$

$$= np\left[q^{n-1} + \frac{2(n-1)}{2} p\, q^{n-2} + \frac{3}{3!}(n-1)(n-2)p^2 q^{n-3} + \ldots \right.$$

$$\left. \ldots\ldots + p^{n-1}\right]$$

$$= np\left[q^{n-1} + (n-1)p\,q^{n-2} + \frac{(n-1)(n-1)}{2} p^2 q^{n-3} + \ldots + p^{n-1}\right]$$

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

$$= np \qquad \text{as } p+q=1.$$

Now from the definition of moment generating fn, we have

$$M_a(t) = e^{-at} M_0(t) \quad \text{where } M_0(t) \text{ is m.g.f about}$$

Origin Now

$$M_0(t) = E(e^{tx}) = \sum_x {}^nC_x\, p^x q^{n-x} e^{tx}$$

$$= \sum_x {}^nC_x (pe^t)^x q^{n-x}$$

$$= (q+pe^t)^n$$

$\therefore M_\mu(t) = $ M.g.f about mean $np = \mu$

$$= e^{-npt} M_0(t)$$

$$= e^{-npt}(q+pe^t)^n$$

$$= (qe^{-pt} + pe^{(1-p)t})^n$$

$$= (qe^{-pt} + pe^{qt})^n$$

Expanding, we get

$$M_\mu(t) = \left\{q\left(1 - pt + \frac{p^2t^2}{2!} - \frac{p^3t^3}{3!} + \frac{p^4t^4}{4!} + \dots\right) + p\left(1 + qt + \frac{q^2t^2}{2!} + \frac{q^3t^3}{3!} + \frac{q^4t^4}{4!} + \dots\right)\right\}^n$$

$$= \left(q - \cancel{qpt} + q\frac{p^2t^2}{2!} - q\frac{p^3t^3}{3!} + q\frac{p^4t^4}{4!} + \dots + p + \cancel{pqt} + p\frac{q^2t^2}{2!} + p\frac{q^3t^3}{3!} + p\frac{q^4t^4}{4!} \dots\right)^n$$

$$= \left[1 + pq\frac{t^2}{2!} + qp(q^2 - p^2)\frac{t^3}{3!} + qp(q^3 + p^3)\frac{t^4}{4!} \dots\right]^n$$

$$= \left[1 + pq\frac{t^2}{2!} + qp(q-p)\frac{t^3}{3!} + qp(p^2 - pq + q^2)\frac{t^4}{4!} \dots\right]^n$$

$$= 1 + npq\frac{t^2}{2!} + npq(q-p)\frac{t^3}{3!} + [npq(p^2 - pq + q^2) + 3n(n-1)p^2q^2]\frac{t^4}{4!} + \dots$$

Since $M_\mu(t) = 1 + t\mu_1 + \frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \dots$

Comparing the coefficients of like powers of t, we get

$\mu_1 = 0$, $\sigma^2 = \mu_2 = npq$, $\mu_3 = npq(q-p)$

$$\mu_4 = npq[1 + 3(n-2)pq]$$

Also $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{n^2p^2q^2(q-p)^2}{n^3p^3q^3} = \dfrac{(q-p)^2}{npq}$

$\beta_2 = \dfrac{\mu_4}{\mu_2^2} = 3 + \dfrac{1-6pq}{npq}$

$\therefore$ Skewness $\gamma_1 = \sqrt{\beta_1} = \dfrac{q-p}{\sqrt{npq}} = \dfrac{1-2p}{\sqrt{npq}}$

Kurtosis $\gamma_2 = \beta_2$

We can see that skewness is positive for $p < \frac{1}{2}$ & negative for $p > \frac{1}{2}$. When $p = \frac{1}{2}$, skewness is zero i.e curve for B.D is symmtrical.

As n, no. of trials, tends to infinity $\beta_1 \to 0$ & $\beta_2 \to 3$.

Since $\mu_2 = \sigma^2 =$ variance $\therefore \sigma^2 = npq$ whereas $\mu =$ mean $=$ Expectation $= np$.

e.g: Find the binomial prob. distribution whose mean is 3 & variance is 2.

Solu: Given that $np = 3$ & $npq = 2$

$\therefore q = \frac{2}{3} \quad \therefore p = \frac{1}{3}$

& $n = 9$

$\therefore$ B.D is $(q+p)^n = \left(\frac{2}{3} + \frac{1}{3}\right)^9$

e.g: The prob. that a bomb dropped from a plane will strike the target is 1/5. If six bombs are dropped find the prob. that (i) exactly two will strike the target (ii) at least two will strike the target.

Solu: $p =$ prob. of striking $= 1/5$

prob of failure = $q = 4/5$

$n = 6$

$\therefore P(X=x) = {}^6C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{6-x}$

(i) Exactly two will strike the target $= P(X=2)$

$$= {}^6C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4 = 0.246$$

(ii) At least two will strike the target $= P(X=2) + P(X=3) + P(X=4) + \dots + P(X=6)$

$$= {}^6C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4 + {}^6C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^3 + {}^6C_4 \left(\frac{1}{5}\right)^4 \left(\frac{4}{5}\right)^2 + {}^6C_5 \left(\frac{1}{5}\right)^5 \left(\frac{4}{5}\right) + {}^6C_6 \left(\frac{1}{5}\right)^6$$

$$= 0.246 + 0.08192 + 0.01536 + 0.001536 + 0.000064$$
$$= 0.344.$$

e.g: Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys. Assume equal probabilities for boys & girls.

Solu: Here $N = 800$; $n = 5$

p = prob of boy $= 1/2$

q = prob of girl $= 1/2$.

$\therefore$ Prob of having x no of boys $= P(X=x) = 800 \times {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$

$\therefore$ Prob of having 3 boys $= 800 \times {}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$

$$= 800 \times \frac{1}{32} \times \frac{5.4}{3!\,2!} = 250$$

Prob of having 5 girls $= 800 \times {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5$

$$= 800 \times \left(\frac{1}{2}\right)^5 = 25$$

Prob of having 2 or 3 boys $= P(X=2) + P(X=3)$

$$= 800 \times {}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 + 250$$

$$= 250 + 250 = 500.$$

e.g.: Criticize the following statement:
The mean of a B.D is 5 & its variance is 9.

Solu: we have $np = 5$ & $npq = 9$

$$\therefore q = \frac{9}{5} = 1.8 > 1$$

Not possible $\therefore$ statement is wrong.

e.g: The following data are the number of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. fit a B.D to this data.

x :	0	1	2	3	4	5	6	7	8	9	10
f :	6	20	28	12	8	6	0	0	0		0

Solu: Here $n = 10$, $N = 80$, $\Sigma f = 80$

$$\therefore \text{Mean} = np = \frac{\Sigma fx}{\Sigma f} = \frac{20(1) + 28(2) + 12(3) + 8(4) + 6(5)}{80}$$

$$np = \frac{174}{80}$$

$$\therefore p = \frac{174}{800} = 0.2175$$

$$\therefore q = 1 - p = 0.7825$$

$\therefore$ B.D is $N(q+p)^n = 80(0.7825 + 0.2175)^{10}$

e.g: Mean of a B.D is 4 & third moment about mean is 1.92. Find $\beta_1, \beta_2, \gamma_1$ & γ_2 for the distribution

Solu: Mean for BD $= np = 4$

& third moment about mean $\mu_3 = npq(q-p) = 1.92$

$$\therefore 4q(q-p) = 1.92$$

$$q(q-p) = 0.48$$

$$q(2q-1) = 0.48$$

$$2q^2 - q - 0.48 = 0$$

$$(2q + 0.6)(q - 0.8) = 0$$

$$q = -\frac{0.6}{2} \quad \text{or} \quad q = 0.8$$

But $q = -\frac{0.6}{2}$ is not possible $\therefore q = 0.8$

$$p = 1 - q = 0.2$$

$$np = 4 \quad \therefore n = 20$$

Now $\mu_2 = npq = 4 \times 0.8 = 3.2$

$$\mu_4 = npq[1 + 3pq(n-2)]$$

$$= 3.2[1 + 3 \times 0.8 \times 0.2 (18)]$$

$$= 30.848$$

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1.92)^2}{(3.2)^3} = 0.1125$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{30.848}{(3.2)^2} = 3.0125$$

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{0.1125} = 0.3354$$

$$\gamma_2 = \beta_2 - 3 = 0.0125.$$

e.g: In a precision bombing attack there is a 50% chance that any one bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better of completely destroying the target?

soln: Here $p = \frac{1}{2} = q$ & $P(x) = {}^nC_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$

$$= {}^nC_x \left(\frac{1}{2}\right)^n$$

Let n be the required no. of bombs. Out of n bombs at least 2 must hit the target in order to destroy it completely $\therefore$

$$P(2) + P(3) + \ldots + P(n) \geq 0.99$$

$\therefore \quad 1-[P(0)+P(1)] \geq 0.99$

$1-0.99 \geq P(0)+P(1)$

$0.01 \geq {}^nC_0\left(\frac{1}{2}\right)^n + {}^nC_1\left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n + n\left(\frac{1}{2}\right)^n$

$\frac{1}{100} \geq \frac{(n+1)}{2^n}$

$2^n \geq (n+1)100$

By hit & trial we get that the least value of n which satisfy this is 11 $\therefore$ 11 bombs should be dropped.

e.g: for a B.D find p if $n=4$ and $P(X=4) = 6P(X=2)$

Solu: $P(X=4) = {}^4C_4\, p^4 q^0$

$P(X=2) = {}^4C_2\, p^2 q^2$

$\therefore \quad {}^4C_4\, p^4 q^0 = 6\,{}^4C_2\, p^2 q^2$

$p^4 = 6 \times 6\, p^2 q^2$

$\Rightarrow \quad p^2 = 36 q^2 = 36(1-p)^2$

$p = 6(1-p)$

$p = 6/7.$

Poisson Distribution— It is a distribution related to the probabilities of events which are extremly rare, but which have a large number of independent opportunities for occurance, e.g - no. of persons born blind per year in a large city. This distribution can be assumed as limiting case of B.D when $n \to \infty$, $p \to 0$ but $np = m$, a fixed constant. Since it has only one parameter m $\therefore$ this is a monomial distribution.

for a B.D

$$P(X=x) = {}^nC_x\, p^x q^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\, p^x (1-p)^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}(np)^x(1-p)^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}(np)^x\left(1-\frac{np}{n}\right)^{n-x}$$

Taking limits as $n\to\infty$, $p\to 0$ so that $np=m$

$$= \frac{1}{x!}(m)^x \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{n-x}$$

$$= \frac{1}{x!}(m)^x e^{-m}$$

Because $\lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{n-x} = \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{n} \cdot \lim_{n\to\infty}\left(1-\frac{m}{n}\right)^{-x}$

$$= \left(e^{-m/n}\right)^n \cdot 1 = e^{-m}$$

$\therefore$ probabilities of $0, 1, 2 \ldots x \ldots$ successes are

$$e^{-m},\ me^{-m},\ \frac{m^2e^{-m}}{2!},\ \frac{m^3e^{-m}}{3!} \ldots \text{ \& so on.}$$

Now $\sum_{x=0}^{\infty}\frac{e^{-m}m^x}{x!} = e^{-m}\sum_{0}^{\infty}\frac{m^x}{x!} = e^{-m}\cdot e^{m} = 1$ as sum of all probabilities.

e.g: Six coins are tossed 6400 times. Using P.D. what is the prob of getting six heads x times.

Solu: Prob of getting all six heads in a throw of six coins $= \frac{1}{2^6} = \frac{1}{64} = p$.

$n = 6400$

$\therefore np = 6400 \times \frac{1}{64} = 100 = m$

$\therefore$ Prob. of getting six heads x times is.

$$\frac{(100)^x e^{-100}}{x!}$$

e.g: Find the probability that at most 5 defective fuses will be found in a box of 200 fuses if experience shows that 2% of such fuses are defective.

Solu: p = prob. of defective fuse is $= 2\% = \frac{2}{100} = \frac{1}{50}$

$n = 200 \therefore m = np = 200 \times \frac{1}{50} = 4.$

$\therefore$ The prob of at most 5 defective fuse is

$$= P(0) + P(1) + P(2) + P(4) + P(5)$$

$$= \frac{e^{-4}(4)^0}{0!} + \frac{e^{-4}(4)^1}{1!} + \frac{e^{-4}(4)^2}{2!} + \dots + \frac{e^{-4}(4)^5}{5!}$$

$$= e^{-4}\left[1 + 4 + 8 + \frac{32}{3} + \frac{32}{3} + \frac{128}{15}\right] = 0.785$$

e.g: A car hire firm has two cars which it hires day by day. The number of demands for a car on each day is distributed as Poisson variate with $m = 1.5$. Calculate the proportion of days on which (i) neither car is used (ii) some demand is refused.

Solu: The prop. of days on which there are x demands for a car is given by

$$P(X=x) = \frac{e^{-1.5}(1.5)^x}{x!}$$

(i) Proportion of days on which neither car is used

$$P(X=0) = \frac{e^{-1.5}(1.5)^0}{0!} = 0.2231$$

(ii) Prop. of days on which some demand is refused

$$P(X>2) = 1 - P(X \le 2)$$
$$= 1 - [P(X=0) + P(X=1) + P(X=2)]$$
$$= 1 - e^{-1.5}\left[1 + 1.5 + \frac{(1.5)^2}{2}\right] = 0.19126.$$

Mean and Variance —

The mean of a Poisson variable is

$$\mu = \sum_{K=1}^{\infty} K \frac{e^{-m} m^K}{K!} = m \sum_{K=1} K \frac{e^{-m} m^{K-1}}{K!}$$

$$= m \sum_{K=1}^{\infty} \frac{e^{-m} m^{K-1}}{(K-1)!}$$

Taking $K-1=j$ $\quad \mu = m \sum_{j=0}^{\infty} \frac{e^{-m} m^j}{j!} = m$

as $\sum_{j=0}^{\infty} \frac{e^{-m} m^j}{j!} = 1$

Now $E(x^2) = \sum_{K=1}^{\infty} K^2 \frac{e^{-m} m^K}{K!} = m \sum_{K=1}^{\infty} K \frac{e^{-m} m^{K-1}}{(K-1)!}$

Taking $K = (K-1)+1$ to obtain

$$E(x^2) = m \sum_{K=1}^{\infty} \frac{(K-1) e^{-m} m^{K-1}}{(K-1)!} + m \sum_{K=1}^{\infty} \frac{e^{-m} m^{K-1}}{(K-1)!}$$

Ist term summation is the mean of X and equal to m and the summation in IInd term is the sum of probabilities & is equal to 1

$\therefore \quad E(x^2) = m \times m + m \times 1 = m^2 + m$

Now $\quad Var(x) = E(x^2) - [E(x)]^2$

$$= m^2 + m - m^2$$

$$= m$$

$\therefore$ Both mean & variance of Poisson distribution are equal to m.

e.g : If x is a Poisson variate such that

$$P(x=2) = 9P(x=4) + 90P(x=6)$$

Find mean μ & β_1 for the distribution solu: we have $P(X=x) = \frac{e^{-m} m^x}{x!}$; where m is the parameter & $m = np$.

Now $P(X=2) = 9(P(X=4)) + 90\, P(X=6)$

$$\frac{\cancel{e^{-m}} m^2}{2!} = 9 \frac{\cancel{e^{-m}} m^4}{4!} + 90 \frac{\cancel{e^{-m}} m^6}{6!}$$

$$\frac{1}{2!} = \frac{9 m^2}{4!} + \frac{m^4}{6!} \times 90$$

$$m^4 + 3m^2 - 4 = 0$$

$$(m^2+4)(m^2-1) = 0$$

$m^2+4=0$ is not possible $\therefore$

$m^2-1=0$ gives $m = \pm 1$

But $m=-1$ not possible $\therefore m=1$.

$\therefore$ mean $\mu = m = 1$

further $\mu_2 = m = 1$ = Variance

$\mu_3 = m = 1$

$\therefore \beta_1 = \frac{1}{m} = 1$.

e.g: Criticize the statement :

The mean of a Poisson distribution is 5, while variance is 16.

solu : Since mean = m = variance for P.D $\therefore$ the statement is wrong.

e.g: Fit a P.D to the following :

x :	0	1	2	3	4
f :	46	38	22	9	1

solu : $\text{Mean} = \frac{\Sigma f x}{\Sigma f} = \frac{38 + 44 + 27 + 4}{116} = \frac{113}{116} = 0.974$.

∴ Theoritical frequency i.e

$$P(X=r) = \frac{e^{-0.974}(0.974)^r}{r!} \times 116$$

Now $P(X=0) = 43.7 \simeq 44$

$P(X=1) = 42.5 \simeq 43$

$P(X=2) = 20.7 \simeq 21$

$P(X=3) = 6.73 \simeq 7$

$P(X=4) = 1.63 \simeq 2$.

e.g: A certain screw making machine produces on average of 2 defective screws out of hundred (100), & packs them in boxes of 500. Find the probability that a box contain 15 defective screw.

Solu: p = prob. of defective screw $= \frac{2}{100}$

$n = 500$

$\therefore m = np = 10$

$$\text{Now } P(X=15) = \frac{e^{-m}(10)^{15}}{15!} = \frac{e^{-10}(10)^{15}}{15!}$$

e.g: A distributor of bean seeds determines from extensive tests that 5% of large batch of seeds will not germinate. He sells in packet of 200 & guarantees 90% germination. Determine the probability that a particular packet will voliate the guarantee.

Solu: If p = prob. of seed not germinating $= \frac{5}{100}$

& $n = 200$ then $m = np = 200 \times \frac{5}{100} = 10$

Prob. that a packet has x seeds which do not germinate $= P(X=x) = \frac{e^{-10}(10)^x}{x!}$.

Prob that a packet violate the guarantee =

$$P(X \geq 10) = \sum_{r=10}^{100} \frac{e^{-10}(10)^r}{r!} = 1 - \sum_{r=0}^{10} \frac{e^{-10}(10)^r}{r!}$$

Geometric & Negative Binomial distribution —

In a series of independent trials with constant probability p of a success, let the random variable X denotes the no. of trials until the first success. Then X is a geometric random variable with parameter $0<p<1$ and

$$f(x) = (1-p)^{x-1} p \qquad x = 1, 2, 3 \ldots$$

The mean of a geometric random variable is

$$\mu = \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

$$= p \sum_{k=1}^{\infty} k q^{k-1} \qquad \text{where } q = 1-p$$

$$= p(1-q)^{-2} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= E[X(X-1) + X] - [E(X)]^2$$

$$= [E(X(X-1))] + E(X) - [E(X)]^2$$

Now $E[X(X-1)] = \sum_{x=1}^{\infty} x(x-1) f(x)$

$$= \sum_{x=1}^{\infty} x(x-1) p(1-p)^{x-1}$$

$$= 2pq \sum_{x=2}^{\infty} \frac{x(x-1)}{2} q^{x-2}$$

$$= 2pq(1-q)^{-3} = \frac{2q}{p^2}$$

$\therefore$ $Var(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{2q+p-(p+q)}{p^2}$

$$= \frac{q}{p^2}$$

eg: If a thief is trying to find the key of a locker, out of a keychain with 5 different keys, what is the probability of the thief succeding in finding the right key in 4th attempt?

solu: Since $p = 1/5$, the req. prob is

$$p(X=4) = \frac{1}{5}\left(1-\frac{1}{5}\right)^3 = 0.1024.$$

eg: If the probability that a target is destroyed on any one shot is 0.5, what is the prob. that it would be destroyed on 6th attempt?

solu: $(0.5)^6$

Negative Binomial distribution.—

A generalization of a geometric distribution in which the random variable is the number of Bernoulli's trials required to obtain r successes results in negative Binomial distribution

In a series of Bernoulli's trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until r successes occurs. Then X is a negative Binomial random variable with parameters $0<p<1$ & $r = 1, 2, 3 \ldots$ and

$$f(x) = {}^{x-1}C_{r-1} (1-p)^{x-r} p^r \qquad x = r, r+1, \ldots$$

Because at least r trials are required for obtaining r successes, the range of x is from r to ∞.

In the special case that $r=1$, a negative binomial distribution converts to geometric distribution.

In this distribution, the number of successes is predetermined & the number of trials is random

In this sense, a negative binomial distribution is considered as the opposite (or negative) of Binomial distribution.

If X is a negative binomial random variable with parameters p & r then

$$\mu = E(X) = \frac{r}{p} \quad \& \quad \sigma^2 = V(X) = \frac{r(1-p)}{p^2}.$$

eg: An item is produced in large numbers. The machine is known to produce 5% defectives. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defectives?

Solu: If 2 defective are to be obtained then it can happen in 2 or more trials. The prob of success is 0.05 in each trial. It is NBD & required prob is

$$= P(x=4) + P(x=5) + P(x=6) + \cdots$$

$$= \sum_{x=4}^{\infty} {}^{x-1}C_{2-1} (0.05)^2 (0.95)^{x-2}$$

$$= 1 - \sum_{x=2}^{3} {}^{x-1}C_{2-1} (0.05)^2 (0.95)^{x-2}$$

$$= 1 - [(0.05)^2 + 2(0.05)^2(0.95)]$$

$$= 0.995$$

Hypergeometric distribution —

A set of N objects contains K objects classified as successes & N−K objects classified as failures. A sample of size n objects is selected randomly (without replacement) from the N objects, where $K \leq N$ & $n \leq N$.

Let the random variable X denotes the number of successes in the sample. Then X is a hypergeometric random variable &

$$f(x) = \frac{{}^{K}C_x \; {}^{N-K}C_{n-x}}{{}^{N}C_n} \qquad x = \max(0, n+K-N) \text{ to } \min(K,n)$$

The expression min(K,n) is used in the definition of range of x because maximum no. of successes that can occur in the sample is the smaller of the sample size, n, and the no of successes available, K.

eg: A batch of parts contains 100 parts from a local supplier of tubing & 200 parts from a supplier of tubing in the next state. If four parts are selected randomly without replacement, what is the probability they are all from the local supplier? What is the probability that two or more parts in the sample are from local supplier?

Solu: Here $N = 300$, $K = 100$ (local supplier as success)

$N-K = 200 \qquad n = 4$

$$P(\text{all from local supplier}) = P(X=4) = \frac{{}^{100}C_4 \times {}^{200}C_0}{{}^{300}C_4}$$

$$= 0.0119$$

$$P(\text{two or more parts from local supplier}) = P(X \geq 2)$$

$$= P(X=2) + P(X=3) + P(X=4)$$

$$= \frac{{}^{100}C_2 \times {}^{200}C_2}{{}^{300}C_4} + \frac{{}^{100}C_3 \times {}^{200}C_1}{{}^{300}C_4} + \frac{{}^{100}C_4 \times {}^{200}C_0}{{}^{300}C_4}$$

$$= 0.298 + 0.098 + 0.0119$$

$$= 0.408$$

· Mean & Variance –

If X is a hypergeometric random variable with parameters N, K & n then

$$\mu = E(X) = np \quad \&$$

$$\sigma^2 = V(X) = np(1-p) \times {}^{N-n}C_{N-1}$$

$$\text{where } p = \frac{K}{N}$$

Here p is interpreted as the proportion of successes in the set of N objects.

Continuous Random Variables

A continuous random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

The probability distribution of a continuous random variable is called as probability density function $f(x)$. For a continuous random variable X, Prob. density fn is a function such that

1) $f(x) \geq 0$

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

3) $P(a \leq X \leq b) = \int_a^b f(x)dx$ = area under $f(x)$ from a to b.

For any x_1 & x_2

$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X \leq x_2)$$
$$= P(x_1 < X < x_2)$$

Cumulative distribution function —

For a continuous random variable, distribution fn is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

for $-\infty < x < \infty$

eg: Let the continuous r.v X denotes the current measured in a thin copper wire in mA. Assume that the range of X is [0, 20mA] & assume that Prob. density fn of X is

$$f(x) = 0.05 \qquad 0 \leq x \leq 20.$$

what is the prob. that a current measurement is less than 10 mA? Find cumulative distribution function.

solu: $P(X<10) = \int_0^{10} f(x)\,dx = \int_0^{10} 0.05\,dx$

$= 0.5$

For cumulative distribution fn.

$F(x) = 0$ for $x<0$

$F(x) = \int_0^x f(u)\,du = 0.05x$ for $0 \le x < 20$

& $F(x) = \int_0^x f(u)\,du = 1$ for $20 \le x$

$\therefore$

$$F(x) = \begin{cases} 0 & x<0 \\ 0.05x & 0 \le x < 20 \\ 1 & 20 \le x \end{cases}$$

The prob. density fn of a continuous random variable can be determined from the cumulative distribution fn by differentiating. $\therefore$ given $F(x)$ (Distribution fn)

Prob. density fn. $= f(x) = \dfrac{dF(x)}{dx}$

eg: If $F(x) = \begin{cases} 0 & x<0 \\ 1-e^{-0.01x} & 0 \le x \end{cases}$; find prob. density fn.

solu: $f(x) = \begin{cases} 0 & x<0 \\ 0.01e^{-0.01x} & 0 \le x \end{cases}$ is the prob density fn.

Mean & variance of a continuous random variable —

Suppose X is a continuous r.v with prob. density fn $f(x)$. Then

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

Variance of $X = Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$

$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

Standard deviation $\sigma = \sqrt{\sigma^2}$

Expected Value of a fn of a continuous random variable–

If X is a continuous random variable with probability density fn $f(x)$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx$$

Continuous Uniform distribution–

A continuous random variable X with probability density function

$$f(x) = \frac{1}{(b-a)} \qquad a \leq x \leq b$$

is a continuous uniform random variable.

The mean of the continuous uniform random variable X is

$$E(X) = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$$

The variance of X is

$$V(X) = \int_a^b \left[x - \left(\frac{a+b}{2}\right)\right]^2 / (b-a) \, dx$$

$$= \left[\frac{1}{b-a} \frac{\left[x - \left(\frac{a+b}{2}\right)\right]^3}{3}\right]_a^b$$

$$= \frac{(b-a)^2}{12}$$

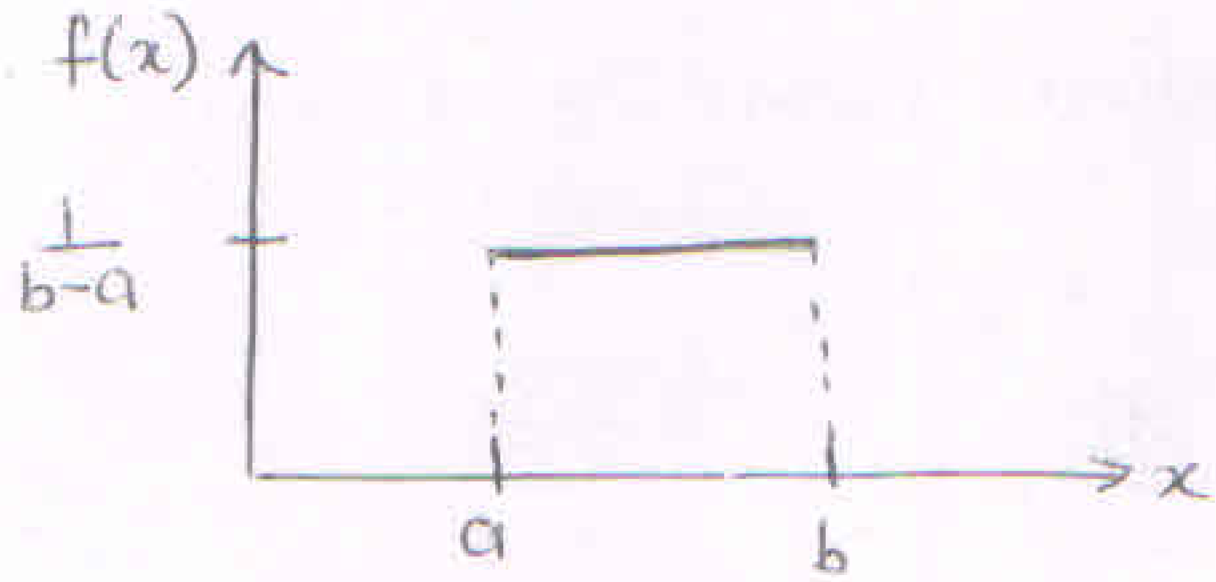
The cumulative distribution function of a continuous uniform random variable is obtained by integration. If $a<x<b$

$$F(x) = \int_a^x \frac{1}{(b-a)} du = \frac{x-a}{b-a}$$

∴ The cumulative distribution of a continuous uniform random variable is

$$F(x)= \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & b \leq x \end{cases}$$

The prob. density fn of a continuous uniform random variable is as follows



eg: Subway trains on a certain line run every half hour between midnight and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes.

Soln: Let the r.v. X denote the waiting time (in minutes) for the next train. Under the assumption that a man arrives at the station at random X. is distributed uniformly on (0, 30) with p.d.f

$$f(x)= \begin{cases} \frac{1}{30} & , 0<x<30 \\ 0 & \text{otherwise} \end{cases}$$

The prob. that he has to wait at least 20 min. is

$$P(X \geqslant 20) = \int_{20}^{30} f(x)\,dx = \frac{1}{30}\int_{20}^{30} 1.\,du = \frac{1}{30}(30-20)$$

$$= \frac{1}{3}$$

Normal distribution—

Any quantity whose variation depends on random causes is distributed according to the normal law. Its importance lies in the fact that a large number of distribution approximate to the normal distribution.

Let us define a variate

$$Z = \frac{x-np}{\sqrt{npq}} = \frac{x-\mu}{\sigma}$$

where x is a binomial variate with mean np and S.D is $\sqrt{npq}$, So that z is a variate with mean zero and variance unity.

$\because$ mean $= E(x)$

$\therefore E(z) = E\left(\frac{x-np}{\sqrt{npq}}\right)$

$= \frac{E(x)}{\sqrt{npq}} - \frac{np}{\sqrt{npq}}$

$= \frac{np}{\sqrt{npq}} - \frac{np}{\sqrt{npq}} = 0$

$\because E(a) = a$

$\because var(ax) = a^2 var(x)$

$\therefore var\left(\frac{x-np}{\sqrt{npq}}\right) = \frac{Var(x)}{npq} - 0$

$= \frac{npq}{npq} = 1$

In the limits as n tends to infinity, the distribution of z becomes a continous distribution extending from $-\infty$ to ∞. The limiting form of B.D

$$Z = \frac{x-np}{\sqrt{npq}}$$

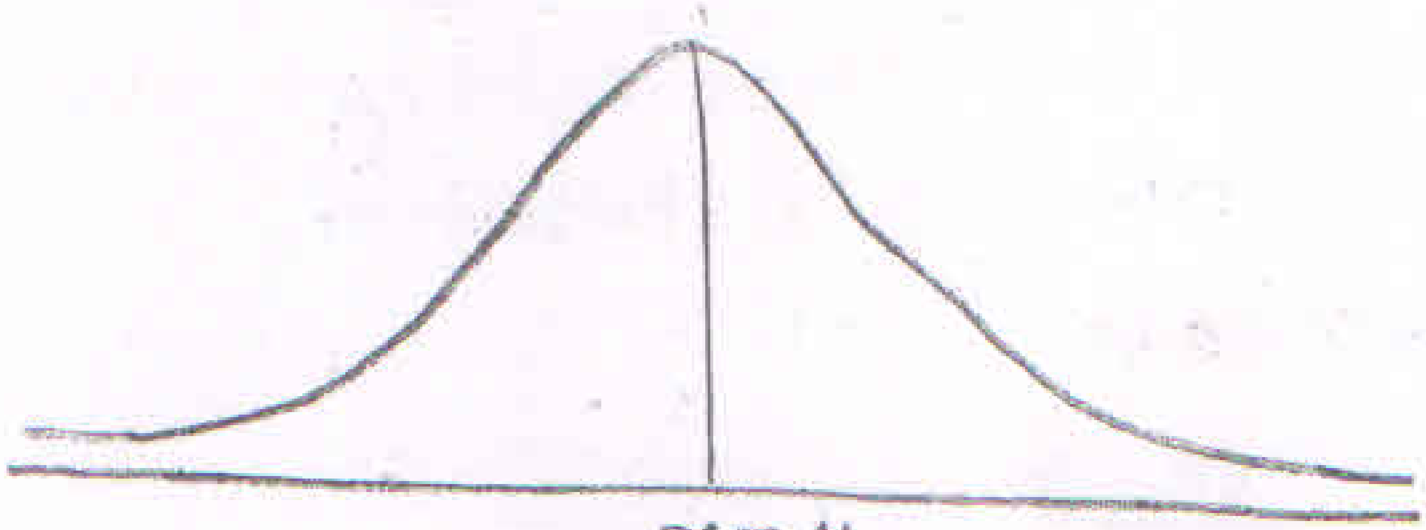
for large values of n, when neither p nor q is very small, is a normal distribution. The normal curve for the dist. would be

Prob. density function

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{——} ①$$

where μ is mean & σ^2 is variance of the distribution.

The curve $y=f(x)$ given by ① is bell shaped and symmetrical about the line $x=\mu$. As x increases numerically, y decreases rapidly & the maximum ordinate is $\frac{1}{\sigma\sqrt{2\pi}}$ which occurs at $x=\mu$. Since the ordinate at $x=\mu$ divides the area under the normal curve into two equal parts, the median of the curve is at $x=\mu$.



Curve never meet X axis i.e $f(x)\neq 0$ for any x. ∵

$$f(x)=0 \quad \text{only if} \quad e^{-(x-\mu)^2/2\sigma^2}=0$$

$$\text{i.e if} \quad \frac{(x-\mu)^2}{2\sigma^2}=\infty \quad \text{i.e } \sigma^2=0 \quad \text{Not possible}$$

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

2) The variate z defined by $z=\frac{x-np}{\sqrt{npq}}$ where x is a B. binomial variate with mean np & variance npq, has its probability density function as

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \quad ; -\infty<z<\infty$$

Standard Normal Variate & Area Under the normal curve —

If X is a normal variate $X \sim N(\mu,\sigma^2)$ then

$$Z=\frac{X-\mu}{\sigma}$$

is called standard normal variate with mean 0 & variance 1 i.e $E(Z)=\text{Expectation}(Z)=0$ &

$$var(Z)=1 \; ∴$$

$$Z \sim N(0,1)$$

The p.d.f of Z is $f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2} \quad -\infty<z<\infty$

The corresponding distribution fn is given by

$$F(z) = P(Z \leq z) = \int_{-\infty}^{z} f(z)\,dz$$

Since $\int_{-\infty}^{\infty} f(z)dz = 1$ & curve is symmetrical about mean i.e $Z=0$ $\therefore$

$$\int_{-\infty}^{0} f(z)dz = \int_{0}^{\infty} f(z)dz = 0.5$$

Also

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 \text{ gives } \int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}$$

Now $P(a \leq Z \leq b) = F(b) - F(a)$

$$= \int_{a}^{b} f(z)dz$$

We write $\phi(z) = \int_{0}^{z} f(z)dz = \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-z^2/2} dz$

The definite integral $\phi(z)$ is called the normal probability integral, which gives the area under the standard normal curve between the ordinates $Z=0$ & $Z=z$. The values of $\phi(z)$ for different values of z is given in table 5 (Appendix-3) of B.S. Grewal.

Since $F(z) = \int_{-\infty}^{z} f(z)dz = \int_{-\infty}^{0} f(z)dz + \int_{0}^{z} f(z)dz$

$$= 0.5 + \phi(z)$$

$$P(a \leq z \leq b) = F(b) - F(a)$$

$$= 0.5 + \phi(b) - 0.5 - \phi(a)$$

$$= \phi(b) - \phi(a)$$

Also $P(-a \leq z \leq a) = P(-a \leq z \leq 0) + P(0 \leq z \leq a)$

$$= P(0 \leq z \leq a) + P(0 \leq z \leq a) \text{ (by symmetry)}$$

$$= 2P(0 \leq z \leq a) = 2\phi(a)$$

e.g: X is a normal variate with mean 30 & s.d 5, find the probabilities that

(i) $26 \le X \le 40$ (ii) $X \ge 45$ (iii) $|X-30| > 5$

Solu: $\mu = 30$ & $\sigma^2 = 25$ $(\sigma = 5)$

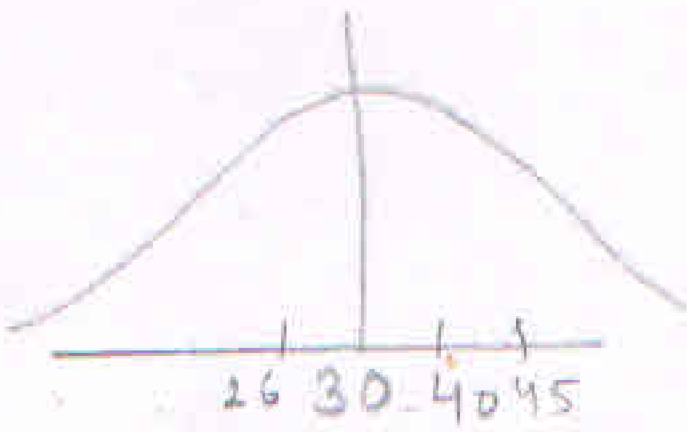
Let $Z = \frac{X-\mu}{\sigma} = \frac{X-30}{5}$, then $Z \sim N(0,1)$.

(i) If $26 \le X \le 40$

$$\frac{26-30}{5} \le Z = \frac{X-30}{5} \le \frac{40-30}{5}$$

$$-\frac{4}{5} \le Z \le 2.$$

$\therefore P(26 \le X \le 40) = P(-0.8 \le Z \le 2)$



$= P(-0.8 \le Z \le 0) + P(0 \le Z \le 2)$

$= P(0 \le Z \le 0.8) + P(0 \le Z \le 2)$

$= 0.2881 + 0.4772$

$= 0.7653$

(ii) $P(X \ge 45) = P\left(Z \ge \frac{45-30}{5}\right)$

$= P(Z \ge 3)$

$= 0.5 - P(0 \le Z \le 3)$

$= 0.5 - 0.4986$

$= 0.00135$

(iii) $P(|X-30| > 5) = P\left(\left|\frac{X-30}{5}\right| > 1\right) = P(|Z| > 1) = 1 - P(|Z| \le 1)$

$= 1 - P(-1 \le Z \le 1) = 1 - 2P(0 \le Z \le 1)$

$= 1 - 2 \times 0.3413 = 1 - 0.6826$

$= 0.3174.$

e.g.: The mean height of 500 students is 151 cm & the S.D is 15 cm. Assuming that the heights are normally distributed, find how many students height lie b/w 120 & 155 cm.

Solu: X = height of student $\sim N(151, 15^2)$.

$$\frac{X-151}{15} = Z \sim N(0,1)$$

$$P(120 \leq X \leq 155) = P\left(\frac{120-151}{15} \leq Z \leq \frac{155-151}{15}\right)$$

$$= P(-2.067 \leq Z \leq 0.26)$$

$$= P(0 \leq Z \leq 2.067) + P(0 \leq 0.26)$$

$$= 0.4803 + 0.1026$$

$$= .5829$$

No of students $= .5829 \times 500 = 291.45$

$= 291$ students

e.g.: In an examination taken by 500 students, the average & S.D of marks obtained (normally distibuted) are 40% & 10%. find approximately

(i) how many will pass, if 50% is fixed as a minimum?

(2) what should be the minimum if 350 candidates are to pass?

(3) How may have scored marks above 60?

Solu: (i) X = marks obtained

$$\frac{X-40}{10} = Z \sim N(0,1)$$

$$P(X \geq 50) = P\left(\frac{X-40}{10} \geq \frac{50-40}{10}\right)$$

$$= P(Z \geq 1) = 0.5 - P(Z \leq 1)$$

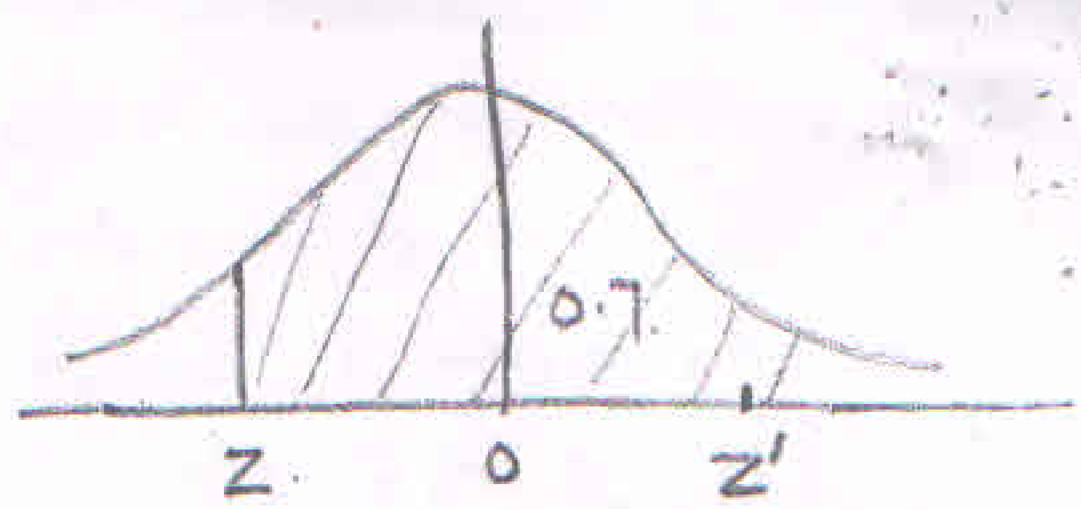
$$= 0.5 - 0.3413 = 0.1587$$

No of students who pass $= 500 \times 0.1587$

$= 79.$

(ii) $P(X \geq \alpha) = \frac{350}{500} = 0.7$

$P\left(Z \geq \frac{\alpha - 40}{10}\right) = 0.7$



Since $P(Z) = 0.7$ $\therefore$ value of Z is negative. Z is symmetrical to Z' where area b/w 0 & Z' is 0.2 $\therefore$

$$P\left(Z \geq \frac{\alpha - 40}{10}\right) = 0.7 \text{ gives}$$

$$\frac{\alpha - 40}{10} = -0.54$$

$$\alpha = -5.4 + 40 = 34.6$$

$\therefore$ $\alpha = 35$ $\therefore$ minimum marks is 35%.

(iii) $P(X \geq 60) = P\left(\frac{X - 40}{10} \geq \frac{60 - 40}{10}\right)$

$= P(Z \geq 2)$

$= 0.5 - P(Z \leq 2)$

$= 0.5 - 0.4772 = 0.0228$

No of students above 60% $= 500 \times 0.0228$
$= 11.$

e.g. Of a large group of men, 5% are under 60 inches in height, & 40% are between 60 & 65 inches. Assuming a normal distribution, find the mean height & a standard deviation.

Solu : If $X \sim N(\mu, \sigma^2)$, then we are given

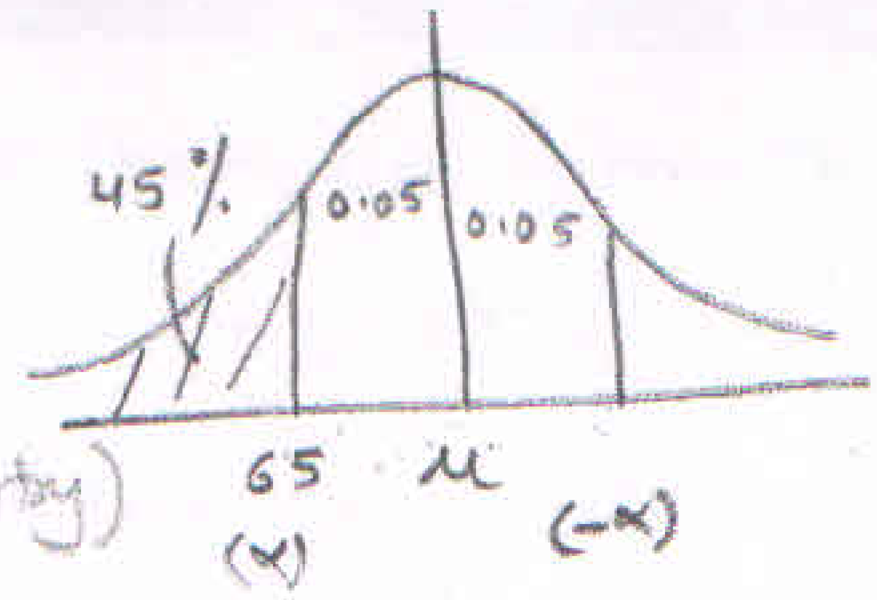
$P(X < 60) = 0.05$

& $P(60 < X < 65) = 0.40$

$P(X < 65) = 0.05 + 0.40 = 0.45$ $\therefore$ $\mu > 65$

$\Rightarrow P(65 < X < \mu) = 0.05$

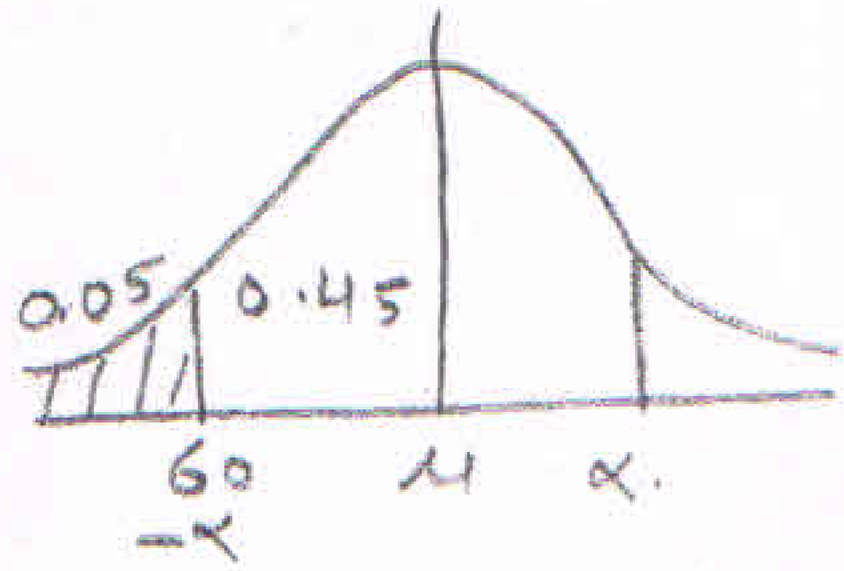
$$P\left(\frac{65-\mu}{\sigma} < \frac{X-\mu}{\sigma} < 0\right) = 0.05$$



$$P\left(0 < Z < \frac{\mu-65}{\sigma}\right) = 0.05 \text{ (by symmetry)}$$

Now $P(X<60) = 0.05$

$$P(60 < X < \mu) = 0.45$$



$$P\left(\frac{60-\mu}{\sigma} < Z < 0\right) = 0.45$$

$$P\left(0 < Z < \frac{\mu-60}{\sigma}\right) = 0.45 \text{ (by symmetry)}.$$

$$\therefore \quad \frac{\mu-65}{\sigma} = 0.13 \quad \& \quad \frac{\mu-60}{\sigma} = 1.64$$

Solving we get $\sigma = 3.29$ &.

$$\mu = 65 + (3.29)(0.13)$$
$$= 65.42.$$

e.g: fit a normal curve to the following data

Length :	8.60	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.52
freq. :	2	3	4	9	10	8	4	1	1

Solu:

$$\text{Mean} = \frac{\Sigma fd}{\Sigma f} = 8.56262 = \mu$$

$$\text{S.D} = \sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$$

$$= 0.0176$$

$$\therefore P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where $\mu = 8.56262$

$\sigma = 0.0176.$

Normal approximation to the Binomial & Poisson distributions—

In case of binomial distribution X with parameters n & p, if n is large (ie $n \to \infty$), the binomial distribution can be approximated to Normal distribution. Keeping in mind that

$$\mu(X) = np \quad \&$$

$$Var(X) = npq = np(1-p)$$

then $$Z = \frac{X-np}{\sqrt{npq}} = \frac{X-np}{\sqrt{np(1-p)}}$$

is approximately a standard normal variate. To approximate a binomial probability with a normal distribution, a continuity correction is applied as follows:

$$P(X \le x) = P(X \le x+0.5)$$
$$= P\left(Z \le \frac{x+0.5-np}{\sqrt{np(1-p)}}\right)$$

&

$$P(x \le X) = P(x-0.5 \le X)$$
$$= P\left(\frac{x-0.5-np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for $np > 5$ & $n(1-p) > 5$.

Ques— The mean yield for one acre plot is 662 Kg with a s.d 22 Kg. How many one-acre plots in a batch of 1000 plots would you expect to have yield over 700 Kg?

Solu: If $X \sim N(\mu, \sigma^2)$ denotes the yields in Kg for one acre plot, then, the prob that a plot yield over 700 Kg is

$$P(X > 700) = P\left(Z > \frac{700-662}{32}\right)$$

$= P(Z > 1.19) = 0.5 - P(0 < Z < 1.19)$

$= 0.5 - 0.3830 = 0.1170.$

Hence in a batch of 1000 plots, the expected no. of plots with yield over 700 Kg is $1000 \times 0.1170 = 117$.

If X is a Poisson random variable with $E(X) = \lambda$ & $Var(X) = \lambda$ then

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

is approximately a standard normal random variate. The same continuity correction used for Binomial distribution can also be applied. The approximation is good for $\lambda > 5$.

eg: In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that the prob. that a bit is received in error is 1×10^{-5}. If 16 million bits are transmitted what is the prob. that 150 or fewer errors occur?

Solu: Let X = no of errors, then

$$P(X \le 150) = \sum_{x=0}^{150} {}^{16\,million}C_x \, (10^{-5})^x (1-10^{-5})^{16 million - x}$$

which is difficult to compute. Using Normal approximation

$$\mu = np = \underbrace{(10^{-5})}_{p} \times \underbrace{(16 \times 10^6)}_{n} = 160$$

$$Var(X) = npq = (16 \times 10^{+6})(10^{-5}) \times (1 - 10^{-5}) = 160(1 - 10^{-5})$$

$$P(X \le 150) = P(X \le 150.5)$$

$$= P\left(Z \le \frac{150.5 - 160}{\sqrt{160(1-10^{-5})}}\right)$$

$$= P(Z \le -0.75) = 0.5 - P(0 \le Z \le 0.75)$$

$$= P(Z \ge 0.75)$$

$$= 0.5 - 0.2734$$

$$= 0.2266$$

eg: Assume that the number of asbestos particles in a squared meter of dust on a surface follows a poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the prob. that 950 or fewer particles are found?

solo: $P(X \leq 950) = \sum_{x=0}^{950} \frac{e^{-1000}(1000)^x}{x!}$

which is difficult to compute. This can be approximated as

$$P(X \leq 950) = P(X \leq 950.5)$$

$$= P\left(\frac{X-1000}{\sqrt{1000}} \leq \frac{950.5-1000}{\sqrt{1000}}\right)$$

$$= P(Z \leq -1.57)$$

$$= P(Z \geq 1.57) = 0.5 - P(0 \leq Z \leq 1.57)$$

$$= 0.5 - 0.4418 = 0.0582$$

Exponential distribution —

A continuous random variable X assuming non-negative values is said to have an exponential distribution with parameter $\lambda > 0$, if its p.d.f is given by

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } 0 \leq x < \infty$$

If a r.v. X has an exponential distribution with parameter λ, then

$$\mu = E(X) = \frac{1}{\lambda} \quad \& \quad \sigma^2 = V(X) = \frac{1}{\lambda^2}$$

Proof: $E(X) = \int_0^\infty x f(x)\,dx = \int_0^x x\lambda e^{-\lambda x}\,dx$

$$= \lambda\left[-\lambda^{-1}x e^{-\lambda x} + \int \lambda^{-1} e^{-\lambda x}\,dx\right]$$

$$= \lambda\left[-\lambda^{-1}x e^{-\lambda x} + \lambda^{-2} e^{-\lambda x}\right]_0^\infty = \lambda\left[0 + \lambda^{-2}\right] = \frac{1}{\lambda}$$

$$Var(x) = E(x^2) - [E(x)]^2$$

Now $E(x^2) = \int_0^{\infty} x^2 f(x)dx = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx$

$$= \frac{2}{\lambda^2}$$

$$Var(x) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

eg: Suppose that the time to failure (in hrs) of fans in a personal computer can be modeled by an exponential distribution with $\lambda = 0.0003$

a) what proportion of the fans will last at least 10,000 hrs?

b) what proportion of the fans will last at most 7000 hrs?

Solu: $\lambda = 0.0003$, X = time to failure

a) $P(X > 10{,}000) = \int_{10{,}000}^{\infty} \lambda e^{-\lambda x} dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_{10000}^{\infty}$

$$= 0.0003 \left[\frac{e^{-0.0003x}}{-0.0003}\right]_{10000}^{\infty}$$

$$= 0 + e^{-0.0003 \times 10000} = e^{-3}$$

$$= 0.04978$$

b) $P(X \leq 7000) = \int_0^{7000} \lambda e^{-\lambda x} = \left[-e^{-\lambda x}\right]_0^{7000}$

$$= -e^{-0.0003 \times 7000} + 1$$

$$= -e^{-2.1} + 1 = -0.1224 + 1 = 0.8776$$

Erlang and Gamma distribution — which represents time until rth event in a Poisson process

A random variable X, having the probability density fn

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} \quad \text{for } x > 0 \text{ and } r = 1, 2, 3 \ldots$$

is called an Erlang random variable with parameter r & λ. The Erlang r.v. with $r=1$ is an exponential random variable.

Gamma function – The gamma fn is

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx \quad \text{for } r>0$$

we know that

$$\Gamma(r) = (r-1)\Gamma(r-1)$$

$$\& \quad \Gamma(r) = (r-1)!$$

Also $\Gamma(1) = 0! = 1$ & $\Gamma(1/2) = \sqrt{\pi}$

Using the result $\Gamma(r) = (r-1)!$ in the definition of the density fn of Erlang distribution, we can write another form of Erlang distribution. (Called as Gamma distribution)

Gamma distribution – The random variable X with p.d.f

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} \quad \text{for } x>0$$

is a gamma random variable with parameters $\lambda>0$ & $r>0$. If r is an integer X has an Erlang distribution. The parameters λ & r are often called the scale & shape parameters, respectively.

Mean & Variance – If X is a gamma r.v. with parameters λ & r, then

$$\mu(X) = E(X) = \frac{r}{\lambda} \quad \& \quad \sigma^2 = Var(X) = \frac{r}{\lambda^2}$$

eg: Calls to a telephone system follow a Poisson distribution with a mean of five calls per min.

a) what are the parameters of the time until tenth call?

b) what is the mean time until 10th call?

c) what is the mean time b/w 9th & 10th call?

d) what is the prob. that exactly four calls occurs within one minute?

Solu: Let X be the time until 10th call. Then

$$\lambda = 5 \text{ call/min} \quad \& \quad r = 10$$

$E(X)$ = mean time until 10th call

$$= \frac{r}{\lambda} = \frac{10}{5} = 2$$

$$Var(X) = \frac{r}{\lambda^2} = (2)\frac{1}{\lambda} = \frac{2}{5} = 0.4$$

Mean time b/w 9th & 10th call = ie mean time b/w two consecutive calls

$$= \frac{\text{Mean time until 10th call}}{10}$$

$$= \frac{2 \text{ min}}{10} = 0.2 \text{ min}$$

Prob that exactly four calls occur within one min

Prob that exactly four calls occur within one min
= Prob that time until 10th call is 2.5 min

$$= P(X = 2.5) = \frac{5^{10}(2.5)^9 e^{-5\times 2.5}}{9!} = 0.3825.$$

Weibull distribution —

Weibull distribution is often used to model the time until failure of many different physical systems. The parameters in distribution provide a great deal of flexibility to model systems in which the number of failures increase with time, decrease with time or remains constant with time.

The random variable X with p.d.f

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\delta}\right)^{\beta}\right) \quad \text{for } x>0$$

is a weibull random variable with scale parameter $\delta>0$ & shape parameter $\beta>0$.

When $\beta=1$, the weibull distribution is identical to exponential distribution.

The cumulative distribution fn is often used to compute probabilities. The following result can be applied.

If X has a weibull distribution with parameters δ & β, then the cumulative distribution fn of X is

$$F(x) = 1 - e^{-\left(\frac{x}{\delta}\right)^{\beta}}$$

Also $\mu = E(x) = \delta \gamma\left(1+\frac{1}{\beta}\right)$ &

$$\sigma^2 = Var(x) = \delta^2 \gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\gamma\left(1+\frac{1}{\beta}\right)\right]^2$$

eg: The time to failure (in hrs) of a bearing in a mechanical shaft is satisfactorily ~~measure~~ modeled as a weibull random ~~as~~ variable with $\beta=2$ & $\delta=5000$ hrs. Determine the mean time until failure. Also determine the prob. that a bearing lasts at least 6000 hrs.

Solu: Here $\beta=2$ & $\delta=5000$ hrs ∴

$$\text{Mean time until failure} = E(x) = \delta\gamma\left(1+\frac{1}{\beta}\right)$$

$$= 5000\,\gamma\left[1+\frac{1}{2}\right] = 5000\times\gamma(2)$$

$$= 5000 X$$

$$P(X>6000) = 1 - P(X<6000) = 1 - F(6000)$$

$$= 1 - \left[1 - e^{-(6000/5000)^2}\right]$$

$$= e^{-36/25} = e^{-1.44} = 0.237$$

Lognormal distribution —

If W is a random variable then $X = \exp(W)$ is also a r.v. An important case occurs when W has a normal distribution. Then $\ln(X) = W$ becomes a normal r.v. In such case, the distribution of X is called lognormal distribution.

Probabilities of X are obtained from the transformation to W, but the range of X is $(0, \infty)$. Suppose that W is normally distributed with mean θ & variance ω^2, then cumulative distribution fn of X is

$$F(x) = P(X \leq x) = P\left[\exp(W) \leq x\right]$$

$$= P\left[W \leq \ln(x)\right]$$

$$= P\left[Z \leq \frac{\ln(x) - \theta}{\omega}\right] \quad \text{for } x > 0$$

where Z is a standard normal variable.

Def: Let W have a normal distribution with mean θ & variance ω^2, then $X = \exp(W)$ is a lognormal random variable with probability density fn

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[\frac{(\ln x - \theta)^2}{2\omega^2}\right] \quad 0 < x < \infty$$

The mean & variance of X are

$$E(X) = e^{\theta + \omega^2/2} \quad \& \quad V(X) = \sigma^2 = e^{2\theta + \omega^2}\left(e^{\omega^2} - 1\right)$$

The parameters of a lognormal distribution are θ & ω^2. The mean & variance of the variable X are fns of these parameters. The lifetime of a product that degrades over time is often modeled by a lognormal random variable. A weibull distribution can also be used for the same, but as lognormal distribution is derived from a simple exponential function of normal random variable, so it is easy to understand & easy to evaluate probabilities eg: The lifetime of a semiconductor laser has a lognormal distribution with $\theta = 10$ hrs & $\omega = 1.5$ hrs. what is the probability that the lifetime exceeds 10,000 hrs? what lifetime is exceeded by 99% of laser?

Solo:

$$P(X > 10{,}000) = 1 - P(X \le 10000)$$
$$= 1 - P[\exp(W) \le 10000]$$
$$= 1 - P[W \le \ln(10000)]$$
$$= 1 - P\left[\frac{W-10}{1.5} \le \frac{\ln(10{,}000)-10}{1.5}\right]$$
$$= 1 - P[Z \le -0.52]$$
$$= 1 - 0.30 = 0.70$$

Now, we have to determine the value of x such that $P(X > x) = 0.99$

$$\therefore \quad P(X > x) = P(\exp(W) > x) = 0.99$$
$$= P(W > \ln(x)) = 0.99$$
$$= 1 - P\left[Z > \frac{\ln(x)-10}{1.5}\right] = 0.99$$
$$\therefore \quad P\left(Z > \frac{\ln(x)-10}{1.5}\right) = 0.01$$
$$\frac{\ln(x)-10}{1.5} = -2.33$$
$$x = \exp(6.505) = 668.48 \text{ hrs.}$$

eg: The lifetime of a semiconductor laser has a lognormal distribution with $\theta = 10$ hrs & $\omega = 1.5$ hrs. what is the probability that the lifetime exceeds 10,000 hrs? what lifetime is exceeded by 99% of laser?

Solo:

$$P(X > 10{,}000) = 1 - P(X \leq 10000)$$

$$= 1 - P[\exp(W) \leq 10000]$$

$$= 1 - P[W \leq \ln(10000)]$$

$$= 1 - P\left[\frac{W-10}{1.5} \leq \frac{\ln(10{,}000)-10}{1.5}\right]$$

$$= 1 - P[Z \leq -0.52]$$

$$= 1 - 0.30 = 0.70$$

Now, we have to determine the value of x such that

$$P(X > x) = 0.99$$

$$\therefore \quad P(X > x) = P(\exp(W) > x) = 0.99$$

$$= P(W > \ln(x)) = 0.99$$

$$= 1 - P\left[Z > \frac{\ln(x)-10}{1.5}\right] = 0.99$$

$$\therefore \quad P\left(Z > \frac{\ln(x)-10}{1.5}\right) = 0.01$$

$$\frac{\ln(x)-10}{1.5} = -2.33$$

$$x = \exp(6.505) = 668.48 \text{ hrs.}$$

Beta distribution –

A continuous distribution that is flexible, but bounded over a finite range is useful for probability models. The random variable X with probability density fn

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } x \in [0,1]$$

is a beta random variable with parameters $\alpha > 0, \beta > 0$

The shape parameters α & β allow the probability density function to assume many different shapes.

If $\alpha = \beta$, the distribution is symmetric about $x = 0.5$ & if $\alpha = \beta = 1$, the beta distribution equals a continuous uniform distribution

In general, there is not a closed form expression for the cumulative distribution function & probabilities for beta random variables need to be computed numerically

If X has a beta distribution with parameters α & β

$$\mu = E(X) = \frac{\alpha}{\alpha+\beta}$$

$$\sigma^2 = Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

eg: Consider the completion time of a large commercial development. The proportion of the maximum allowed time to complete a task is modeled as a beta random variable with $\alpha = 2.5$ & $\beta = 1$. What is the probability that the proportion of the maximum time exceeds 0.7?

Solu: Suppose X denotes the proportion of the maximum time required to complete the task. The prob is

$$P(X > 0.7) = \int_{0.7}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \int_{0.7}^{1} \frac{\Gamma(3.5)}{\Gamma(1)\Gamma(2.5)} x^{1.5}\,dx$$

$$= \left[\frac{(2.5)(1.5)(0.5)\sqrt{\pi}}{1.(1.5)(0.5)\sqrt{\pi}} \frac{x^{2.5}}{2.5}\right]_{0.7}^{1}$$

$$= 1 - (0.7)^{2.5} = 0.59$$