UNIT- I

PROBABILITY AND RANDOM VARIABLES

AXIOMS OF PROBABILITY

(i) $0 \le P(E) \le 1$ (ii) $P(S) = 1$.

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Thm: 1 The probability of an impossible event is zero (or) The null event has probability 0 ie $P(\phi) = 0$

Pf If we consider a sequence of events $E_1, E_2, \ldots$ where $E_1 = S$, $E_i = \phi$ for $i > 1$, then the events are mutually exclusive and as $S = \bigcup_{i=1}^{\infty} E_i$

$$P(S) = \sum_{i=1}^{\infty} P(E_i)$$

$$= P(E_1) + \sum_{i=2}^{\infty} P[E_i]$$

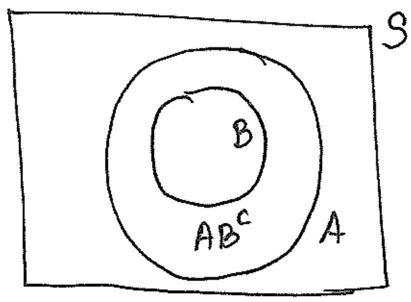
$$= P(S) + \sum_{i=2}^{\infty} P(\phi)$$

$$\sum_{i=2}^{\infty} P(\phi) = 0$$

ie, $P(\phi) = 0$

Thm: 2 If $B \subset A$; $P(B) \leq P(A)$

Pf



B and AB^c are mutually exclusive events such that

$B \cup AB^c = A$

Note: (i) $AB^c = A \cap B^c$

(ii) $\overline{B} = B^c$

$P(B \cup AB^c) = P(A)$

$P(B) + P(AB^c) = P(A)$

$P(B) \leq P(A)$.

Note: $\overline{A} = A^c$.

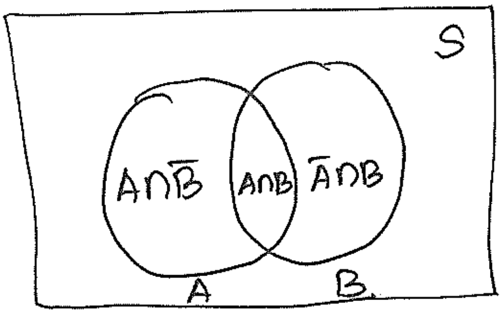
Thm: 4 Additional law of Probability

If A and B are any two events, and are not disjoint then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

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

(B)

Pf



we get the events A and $\bar{A}\cap B$ are disjoint

$$A\cup B = A\cup(\bar{A}\cap B)$$

$$P(A\cup B) = P[A\cup(\bar{A}\cap B)]$$

$$= P(A) + P(\bar{A}\cap B)$$

adding and subtracting $P(A\cap B)$ we get

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

$$= P(A) + P[(\bar{A}\cap B)\cup(A\cap B)] - P(A\cap B)]$$

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

$\because (\bar{A}\cap B)\cup(A\cap B) = B$

Problems

(1) If two dice are rolled, what is the probability that the sum of the upturned faces will be equal to 7?

Sol

$$n(S) = 36$$

Let $A = \{$ Sum of the topturned faces will equal 7 $\}$

$= \{ (1,6), (2,5)\ (3,4), (4,3), (5,2), (6,1) \}$

$n(A) = 6$

$$P(A) = \frac{n(A)}{n(S)} = \frac{6}{36} = \frac{1}{6}.$$

(2) A bag contains 5 white and 10 red balls. Three balls are taken out at random. Find the probability that all the three balls drawn red.

Sol

Total number of balls $= 15$

$S = \{$ Three balls are taken out of 15 $\}$

$$n(S) = {}^{15}C_3 = \frac{15 \cdot 14 \cdot 13}{3 \cdot 2 \cdot 1} = 455$$

Number of red balls $= 10$

$A = \{$ Three balls which are red $\}$

$$n(A) = {}^{10}C_3 = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{120}{455} = \frac{24}{91}$$

(3) A lot of integrated circuit chips consists of 10 good, 4 with minor defects and 2 with major defects. Two chips are randomly chosen from the lot. What is the probability that atleast one chip is good? [M/J - 2017].

Sol

$$P(\text{atleast one is good}) = \frac{n(A)}{n(S)} = \frac{(^{10}C_1)(^{6}C_1) + {^{10}C_2}}{^{16}C_2}$$

$$= \frac{(10)(6)+45}{120}$$

$$= \frac{60+45}{120} = \frac{105}{120}$$

$$= \frac{7}{8}$$

(4) Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. Show that the chance that exactly two of them will be children is $\frac{10}{21}$

Sol

Total no. of persons $= 9$

4 persons can be chosen out of 9 persons $= {^9C_4}$ ways

$$= \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1}$$

$$= 126 \text{ ways.}$$

2 children out of 4 children $= 4_{C_2}$ ways

$= \frac{4 \cdot 3}{2 \cdot 1} = 6$ ways

The remaining two persons can be choosen from

5 person (3 men + 2 women) $= 5_{C_2}$ ways

$= \frac{5 \times 4}{2 \times 1} = 10$ ways

$\therefore$ The number of favourable case $= 4_{C_2} \times 5_{C_2}$ ways

$= 6 \times 10$

$= 60$ ways

$\therefore$ Required probability $= \frac{60}{126} = \frac{10}{21}$.

Type-I Mutually exclusive events (disjoint)

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

(1) One card is drawn from a pack of 52 cards. What is the probability that it is either a king or a queen.

D

$A = \{$an event that the card drawn is King$\}$

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

$B = \{$an event that the card drawn is queen$\}$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

$A \cup B = \{$an event that the card to be either a King or a queen$\}$

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

$$= \frac{1}{13} + \frac{1}{13} = \frac{2}{13}$$

(2) A bag contains 30 balls numbered from 1 to 30. one 30. one ball is drawn at random. Find the probability that the number of the ball drawn will be a multiple of (a) 5 or 7 and (b) 3 or 7

Sol

Given: $n(S) = 30$

let A = The Probability of the number being multiple of 5

$$P(A) = P(5, 10, 15, 20, 25, 30) = \frac{6}{30}$$

let B = The Probability of the number being multiple of 7.

$P(B) = P(7, 14, 21, 28) = \frac{4}{30}$

let C = The probability of the number being multiple of 3

$P(C) = P(3, 6, 9, 12, 15, 18, 21, 24, 27, 30) = \frac{10}{30}$

(a) The events A and B are mutually exclusive that. Probability of the number being a multiple of 5 or 7 will be

$$= \frac{6}{30} + \frac{4}{30} = \frac{10}{30}$$

(b) The events C and B are not mutually exclusive

$$P(C \cap B) = P[21] = \frac{1}{30}$$

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

$$= \frac{10}{30} + \frac{4}{30} - \frac{1}{30} = \frac{13}{30}$$

* Not Mutually exclusive, independent events

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

(ii) $P(A \cap B) = P(A) \cdot P(B)$.

1) A can hit target in 4 out of 5 shots and B can hit the target in 3 out of 4 shots. Find the probability that (i) the target being hit when both try (ii) the target being hit by exactly one person.

E

(b) Let A, B the events

A hit the target $P(A) = \frac{4}{5}$

B hit the target $P(B) = \frac{3}{4}$

(i) The events A and B are not mutually exclusive because both of them hit the target.

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

$= \frac{4}{5} + \frac{3}{4} - P(A) \cdot P(B)$

$= \frac{4}{5} + \frac{3}{4} - \left(\frac{4}{5}\right)\left(\frac{3}{4}\right)$

$= \frac{16+15}{20} - \frac{12}{20} = \frac{31}{20} - \frac{12}{20} = \frac{19}{20}$

$= \frac{19}{20}$

(ii) The target being hit by exactly one person.

$= P[(A \cap \bar{B}) \cup (B \cap \bar{A})]$

$= P[(A \cap \bar{B}) + (B \cap \bar{A})]$

$= P(A) P(\bar{B}) + P(B) \cdot P(\bar{A})$

$= P(A)[1 - P(B)] + P(B)[1 - P(A)]$

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

$$= P(A) + P(B) - 2P(A) \cdot P(B)$$

$$= \frac{4}{5} + \frac{3}{4} - 2\left(\frac{4}{5}\right)\left(\frac{3}{4}\right)$$

$$= \frac{16+15-24}{20}$$

$$= \frac{7}{20} \text{ //}$$

(2) One card is drawn from a deck of 52 cards. What is the probability of the card being either red or a king.

Sol let $A = \{$an event that the card drawn is red$\}$

$B = \{$an event that the card is King$\}$

$A \cup B = \{$an event that a card to be either red or a king$\}$

$$P(A) = \frac{n(A)}{n(S)} = \frac{26}{52} = \frac{1}{2}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{4}{52} = \frac{1}{13}$$

There are two red coloured king cards

$$P(A \cap B) = \frac{2}{52} = \frac{1}{26}.$$

(F)

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

$$= \frac{1}{2}+\frac{1}{13}-\frac{1}{26}$$

$$= \frac{13+2-1}{26}$$

$$= \frac{14}{26}$$

$$= \frac{7}{13}$$ //

(3) If A and B are events with $P(A)=\frac{3}{8}$, $P(B)=\frac{1}{2}$ and $P(A\cap B)=\frac{1}{4}$, find $P(A^c\cap B^c)$.

Sol

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

$$P(A\cup B) = \frac{3}{8}+\frac{1}{2}-\frac{1}{4} = \frac{3+4-2}{8} = \frac{5}{8}.$$

$$P(A^c\cap B^c) = 1-P(A\cup B) = 1-\frac{5}{8} = \frac{3}{8}$$ //

(4) If A and B are independent events with $P(A)=0.4$ & $P(B)=0.5$ find $P(A\cup B)$.

Sol $$P(A\cup B) = P(A)+P(B)-P(A\cap B) = 0.4+0.5-(0.4)(0.5)$$

$$= 0.9-0.2$$

$$= 0.7$$ //

CONDITIONAL PROBABILITY

The conditional probability of A given B is

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$ if $P(B) \neq 0$ and it is undefined otherwise

(1) A bag contains 5 red and 3 green balls and a second bag 4 red and 5 green balls. one of the bages is selected at random and a draw of 2 balls is made from it. what is the probability that one of them is red and the other is green.

Soln

let A_1 and A_2 denote the event of selecting the first bag and second bag resp.

$P(A_1) = \frac{1}{2} = P(A_2)$ and A_1 and A_2 are mutually exclusive events.

$S = A_1 \cup A_2$

let B denote the event of selecting one red and one green ball.

$$P(B/A_1) = \frac{5C_1 \times 3C_1}{8C_2} = \frac{5\times 3}{\frac{8\times 7}{2\times 1}} = \frac{15}{56}\times 2 = \frac{15}{28}$$

$$P(B/A_2) = \frac{4C_1 \times 5C_1}{9C_2} = \frac{4\times 5}{\frac{9\times 8}{2\times 1}} = \frac{20}{72}\times 2 = \frac{5}{9}$$

$\therefore$ The required probability $= P(A_1) * P(B/A_1) + P(A_2) \cdot P(B/A_2)$

$$= \left(\frac{1}{2}\right)\left(\frac{15}{28}\right) + \frac{1}{2}\left(\frac{5}{9}\right)$$

$$= \frac{15}{56} + \frac{5}{18}$$

$$= \frac{275}{504} //$$

(2) A box contains 4 bad and 6 good tubes. Two are drawn out from the box at a time. One of them is tested and found to be good. What is the probability that the ~~order~~ other one is also good?

Sol: Let A = one of the tubes drawn is good

B = the other tube is good.

$P(A \cap B) = P$[both the tubes drawn are good]

$$= \frac{6C_2}{10C_2} = \frac{\frac{6\times5}{1\times2}}{\frac{10\times9}{2\times1}} = \frac{6\times5}{10\times9} = \frac{1}{3}$$

Knowing that one tube is good, the conditional probability that the other tube is also good.

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

$$= \frac{1/3}{(6/10)} = \left(\frac{1}{3}\right)\left(\frac{10}{6}\right) = \frac{5}{9}.$$

(3) In a certain group of computer personnel, 65% have insufficient knowledge of hardware, 45% have inadequate idea of software and 70% are in either one or both of the two categories. What is the percentage of people who know software among those who have a sufficient knowledge of hardware?

Sol

Let $P(A)$ = probability of people having knowledge insufficient knowledge of hardware

$$= 65\% = \frac{65}{100} = 0.65$$

$$P(\bar{A}) = 1 - P(A) = 1 - 0.65 = 0.35$$

$P(B)$ = probability of people having inadequate idea of software

$$= 45\% = \frac{45}{100} = 0.45$$

$$P(\bar{B}) = 1 - P(B) = 1 - 0.45 = 0.55$$

$P(A \cup B) = 70\%$ [either one or both]

$$= \frac{70}{100} = 0.70$$

$$P(\overline{A \cup B}) = 1 - P(A \cup B)$$
$$= 1 - 0.70$$
$$= 0.30$$

$$P(\bar{B}/\bar{A}) = \frac{P(\overline{A \cup B})}{P(A)} = \frac{0.30}{0.35} = 0.8570.$$

(4) An urn contains 10 white and 3 black balls. Another urn contains 3 white and 5 black balls. Two balls are drawn at random from the first urn and placed in the second urn and then 1 ball is taken at random from the latter. What is the probability that it is a white ball?

Sol

Let B_1 = event of drawing 2 white balls from the first urn

B_2 = event of drawing 2 black balls from it and

B_3 = event of drawing 1 white and 1 black ball from it

Clearly $B_1, B_2 \& B_3$ are mutually exclusive and exhaustive events

Let A = event of drawing a white ball from the second urn after transfer

$$P(B_1) = \frac{10_{C_2}}{13_{C_2}} = \frac{10 \times 9}{2 \times 1} \times \frac{2 \times 1}{13 \times 12} = \frac{15}{26}$$

$$P(B_2) = \frac{3C_2}{13C_2} = \frac{1}{26}$$

$$P(B_3) = \frac{10C_3}{13C_2} = \frac{10}{26}$$

$P\left(\frac{A}{B_1}\right)$ = P(drawing a white ball / 2 white balls have been transferred).

= P(drawing a white balls / urn II contains 5 white and 5 black balls)

$$= \frac{5}{10}$$

Ily

$$P\left(\frac{A}{B_2}\right) = \frac{3}{10}$$

$$P\left(\frac{A}{B_3}\right) = \frac{4}{10}.$$

by theorem of total probability

$$P(A) = P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2) + P(B_3) P(A/B_3)$$

$$= \left(\frac{15}{26}\right)\left(\frac{5}{10}\right) + \left(\frac{1}{26}\right)\left(\frac{3}{10}\right) + \left(\frac{10}{26}\right)\left(\frac{4}{10}\right)$$

$$= \frac{59}{130} \text{ //.}$$

BAYE'S THEOREM

Baye's theorem or Theorem of Probability of cases

let $B_1, B_2, \ldots B_n$ be an exhaustive and mutually exclusive random experiments and A be an event related to that B_i then

$$P(B_i/A) = \frac{P(B_i)\,P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\,P(A/B_i)}$$

Pf

$P(A \cap B_i) = P(B_i)\,P(A/B_i)$ by conditional probability

$$P(B_i \cap A) = P(A) \cdot P(B_i/A) = P(B_i)\,P(A/B_i)$$

$$P(B_i/A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)\,P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\,P(A/B_i)}$$

$$\boxed{P(B_i/A) = \frac{P(B_i)\,P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\,P(A/B_i)}}$$

(1) The contents of urns I, II, III are as follows

Balls / urns	White	Black	Red
I	1	2	3
II	2	1	1
III	4	5	3

one urn is choose at random and two balls are drawn. They happen to be white and red. What is the probability that they come from urns I, II and III? [M/J-2006, A/M-2008]

Sol) Let B_1, B_2, B_3 denote events that the urns I, II, III are choosen. Let A be the event that the two balls taken from the selected urn are white and red.

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$

$$P(A/B_1) = \frac{{}^1C_1 \times {}^3C_1}{{}^6C_2} = \frac{1 \times 3}{\frac{6 \times 5}{2 \times 1}} = \frac{3}{30} \times 2 = \frac{1}{5}.$$

$$P(A/B_2) = \frac{{}^2C_1 \times {}^1C_1}{{}^4C_2} = \frac{2 \times 1}{4 \times 3} \times 2 \times 1 = \frac{1}{3}$$

$$P(A/B_3) = \frac{{}^4C_1 \times {}^3C_1}{{}^{12}C_2} = \frac{2}{11}.$$

by Baye's Theorem

$$P(B_i/A) = \frac{P(B_i)\,P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\,P(A/B_i)}$$

$$P(B_2/A) = \frac{P(B_2)\,P(A/B_2)}{\sum_{i=1}^{3} P(B_i)\,P(A/B_i)}$$

$$= \frac{P(B_2)\,P(A/B_2)}{P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2) + P(B_3)\,P(A/B_3)}$$

$$= \frac{(1/3)(1/3)}{\left(\frac{1}{3}\right)\left(\frac{1}{5}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{2}{11}\right)}$$

$$= \frac{55}{118}.$$

$$P(B_3/A) = \frac{P(B_3)\,P(A/B_3)}{P(B_1)\,P(A/B_1) + P(B_2)\,P(A/B_2) + P(B_3)\,P(A/B_3)}$$

$$= \frac{(1/3)\left(\frac{2}{11}\right)}{\left(\frac{1}{3}\right)\left(\frac{1}{5}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{2}{11}\right)} = \frac{30}{118}.$$

$$P(B_1/A) = \frac{P(B_1)P(A/B_1)}{P(B_1)P(A/B_1)+P(B_2)P(A/B_2)+P(B_3)P(A/B_3)}$$

$$= 1 - P(B_2/A) - P(B_3-A)$$

$$= 1 - \frac{55}{118} - \frac{30}{118} = \frac{118-85}{118} = \frac{33}{118}$$

(2) A bag A contains 2 white and 3 red balls and a bag B contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and is found to be red. Find the Probability that it was drawn from bag B. [N/D-2006]

Sol

Bag	White	Red
A (B_1)	2	3
B (B_2)	4	5

Let B_1 the event that the ball is drawn from the bag A

B_2 the event that the ball is drawn from the bag B

A be the event that the drawn ball is red.

$P(B_1) = P(B_2) = \frac{1}{2}$

$P(A/B_1) = P(B_2) = 1/2$

$$P(A/B_1) = \frac{3c_1}{5c_1} = \frac{3}{5}, \quad P(A/B_2) = \frac{5c_1}{9c_1} = \frac{5}{9}$$

$$P(B_2/A) = \frac{P(B_2) \cdot P(A/B_2)}{P(B_1)P(A/B_1) + P(B_2)P(A/B_2)}$$

$$\frac{(1/2)(5/9)}{(\frac{1}{2})(\frac{3}{5}) + (\frac{1}{2})(\frac{5}{9})} = \frac{5/18}{52/90} = \frac{25}{52}$$

(3) The members of a consulting firm rent cars from rental agencies. A, B and C as 60%, 30% and 10% respectively. If 9, 20 and 6% of cars from A, B and C agencies need turn up (a) If a rental car delivered to the firm does not need turn up, what is the probability that it came from B agency. (b) if a rental car delivered to the firm need turn up what is the probability that came from B agency. [A/M-2004, 2008]

Sol Let E_1 be the event that the members of agency A

Let E_2 〃 B

Let E_3 〃 C

$$P(E_1) = 60\% = \frac{60}{100} = 0.60$$

$$P(E_2) = 30\% = \frac{30}{100} \; 0.30$$

$$P(E_3) = 10\% = \frac{10}{100} = 0.10$$

Let D be the event that cars need turn up

Let $\bar{D}$ be the event that cars need not turn up.

$$P(D/E_1) = 9\% = 9/100 = 0.09$$

$$P(D/E_2) = 20\% = \frac{20}{100} = 0.20$$

$$P(D/E_3) = 6\% = \frac{6}{100} = 0.06$$

$$P(\bar{D}/E_1) = 1 - P(D/E_1) = 1 - 0.09 = 0.91$$

$$P(\bar{D}/E_2) = 1 - P(D/E_2) = 1 - 0.20 = 0.80$$

$$P(\bar{D}/E_3) = 1 - P(D/E_3) = 1 - 0.06 = 0.94$$

(a) To find $P(E_2/\bar{D})$

$$P(E_2/\bar{D}) = \frac{P(E_2) \cdot P(\bar{D}/E_2)}{P(E_1)P(\bar{D}/E_1) + P(E_2)P(\bar{D}/E_2) + P(E_3)P(\bar{D}/E_3)}$$

$$= \frac{(0.3)(0.8)}{(0.6)(0.91) + (0.3)(0.8) + (0.1)(0.94)}$$

$$\frac{0.24}{0.546 + 0.24 + 0.094} = \frac{0.24}{0.88} = 0.2727$$

(b) To find $P(E_2/D)$

$$P(E_2/D) = \frac{P(E_2)P(D/E_2)}{P(E_1)P(D/E_1) + P(E_2)P(D/E_2) + P(E_3)P(D/E_3)}$$

$$\frac{(0.30)(0.20)}{(0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06)} = \frac{1}{2}$$

Random Variablee:

A Real variable x whose values is determined by the output of the Random experiments is called a Random variables.

Eg: A Random experiment consists of two tosses of a coin.
Consider the Random variables which of the number of heads.

head {0,1,2}.

Outcome	HH	HT	TH	TT
Value of x	2	1	1	0

Type of Random variables:

i) Discrete Random Variable
ii) Continuous Random Variable.

Discrete Random Variable:

The Random Variable which can assume only a countable number of Real values is called Discrete Random Variable.

eg: ① No. of Telephone calls per unit time
② No. of Printing mistakes in each page of a Book.

Continuous Random Variable:

A Random Variable x is said to be continuous if it can take all possible values between certain limits.
Ex: The time that you spend for studies during a day

Probability Mass Function: (P.m.f) (Discrete)

i) $P(x_i) \geq 0$

ii) $\sum_{i=0}^{\infty} P(x_i) = 1.$

Probability Density Function: (continuous) (pdf)

i) $f(x) > 0$.

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

iii) $P(x=a) = \int_a^a f(x)\, dx = 0.$

iv) $P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)\, dx$

Probability Distribution Function: [F(x)] [PDF] [Discrete case].

$F(x) = P(X \leq x)$; $-\infty < x < \infty$

Cumulative Distribution Function (Continuous case)

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

Note:

$P(a < x < b) = F(b) - F(a).$

Relation pdf & PDF:

p.d.f $f(x) = \frac{d}{dx}[F(x)] = F'(x)$

Conditional Probability:

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

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

1. A Discrete Random variable X has the following 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) Find the minimum of 'a' such that $P(X \leq a) > 1/2$

iv) Determine the Distribution of X

v) Evaluate $P(1.5 < X < 4.5 / X > 2)$

i) W.K.T; $\sum_{i=0}^{7} P(x_i) = 1$

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

$$10K^2 + 9K - 1 = 0.$$

$$(10K - 1)(K + 1) = 0$$

$\boxed{K = \frac{1}{10}}$ (or) $\boxed{K = -1}$

$K = -1$ is not possible.

$\therefore \boxed{K = \frac{1}{10}}$

Probability Function:

X	0	1	2	3	4	5	6	7
$P(x)$	0	$1/10$	$2/10$	$2/10$	$3/10$	$\frac{1}{100}$	$\frac{2}{100}$	$\frac{17}{100}$

ii) $P(X<6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

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

$$= \frac{8}{10} + \frac{1}{100} = \frac{81}{100}.$$

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

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

$P(0 < X < 5) = P(X=1) + P(X=2) + P(X=2) + P(X=3) + P(X=4)$

$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} = \frac{8}{10}$

(iii) $P(X \le a) > 1/2$

$P(X \le 0) = 0$

$P(X \le 1) = 1/10$

$P(X \le 2) = 3/10$

$P(X \le 3) = 5/10$

$P(X \le 4) = 8/10 = \frac{4}{5} = 0.8 > 1/2$

$\boxed{a = 4}$

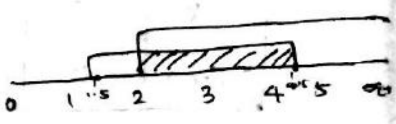
iv) Probability Distribution Function.

x	$P(x)$	$F(x) = P(X \le x)$
0	0	0
1	1/10	1/10
2	2/10	3/10
3	2/10	5/10
4	3/10	8/10
5	1/100	81/100
6	2/100	83/100
7	17/100	1

v) $P(1.5 < X < 4.5 \,/\, X > 2)$

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

$= \frac{P(1.5 < X < 4.5 \cap X > 2)}{P(X > 2)}$



$= \frac{P(2 < X < 4.5)}{1 - P(X \le 2)}$

$= \frac{P(3) + P(4)}{1 - \frac{3}{10}} = \frac{\frac{2}{10} + \frac{3}{10}}{\frac{7}{10}} = \frac{\frac{5}{10}}{\frac{7}{10}} = \boxed{\frac{5}{7}}$

2. A discrete Random Variable X has the following Probability Distribution.

$X=x$	0	1	2	3	4	5	6	7	8
$P(X=x)$	a	3a	5a	7a	9a	11a	13a	15a	17a

i) Find the value of 'a'.

ii) Find $P(X<3)$, $P(0<X<3)$, $P(X\geq 3)$.

iii) Find the Distribution function of X.

i) $$\sum_{i=0}^{7} P(x_i) = 1$$

$$a+3a+5a+7a+9a+11a+13a+15a+17a-1=0.$$

$$81a-1=0$$

$$\boxed{a = 1/81}$$

ii) $P(X<3) = P(X=0)+P(X=1)+P(X=2)$

$=$

Probability function:

$X=x$	0	1	2	3	4	5	6	7	8
$P(X=x)$	$1/81$	$3/81$	$5/81$	$7/81$	$9/81$	$11/81$	$13/81$	$15/81$	$17/81$

ii) $P(X<3) = P(X=0)+P(X=1)+P(X=2)$

$$= \frac{1}{81}+\frac{3}{81}+\frac{5}{81}$$

$$= \frac{9}{81} = \boxed{\frac{1}{9} = P(X<3)}$$

$P(0<X<3) = P(X=1)+P(X=2)$

$$= \frac{3}{81}+\frac{5}{81} = \boxed{\frac{8}{81} = P(0<X<3)}$$

$P(X\geq 3) = 1-P(X<3)$

$$= 1-[P(X=0)+P(X=1)+P(X=2)]$$

$$= 1-\left[\frac{1}{81}+\frac{3}{81}+\frac{5}{81}\right]$$

$$= 1-\left[\frac{9}{81}\right] \Rightarrow \frac{81-9}{81} = \boxed{\frac{72}{81} = P(X\geq 3)}$$

iii) Probability Distribution function:

x	$P(x)$	$F(x) = P(X \le x)$
0	1/81	1/81
1	3/81	4/81
2	5/81	9/81
3	7/81	16/81
4	9/81	25/81
5	11/81	36/81
6	13/81	49/81
7	15/81	64/81
8	17/81	1

(3) A Random Variable x has following Probability Distribution.

$X=x$	-2	-1	0	1	2	3
$P(X=x)$	0.1	k	0.2	2k	0.3	3k

i) Find the value of k?

ii) Evaluate $P(X<2)$, $P(-2<X<2)$

iii) Find the cumulative Distribution.

i) $$\sum_{i=-2}^{3} P(x_i) = 1.$$

$$0.1 + k + 0.2 + 2k + 0.3 + 3k - 1 = 0.$$

$$+0.6 + 6k - 1 = 0.$$

$$6k = 1 - 0.6$$

$$k = \frac{0.4}{6} = 0.07. \quad \boxed{k = 0.07}$$

Probability Function:

$X=x$	-2	-1	0	1	2	3
$P(X=x)$	0.1	0.07	0.2	0.14	0.3	0.21

ii) $P(x<2) = P(x=-2) + P(x=-1) + P(x=0) + P(x=1)$

$= 0.1 + 0.07 + 0.2 + 0.14$

$\boxed{P(x<2) = 0.51}$

$P(-2<x<2) = P(x=-1) + P(x=0) + P(x=1)$

$= 0.07 + 0.2 + 0.14$

$\boxed{P(-2<x<2) = 0.41}$

iii) cumulative distribution / Probability

x	$P(x)$	$F(x) = P(X \le x)$
-2	0.1	0.1
-1	0.07	0.17
0	0.2	0.37
1	0.14	0.51
2	0.3	0.81
3	0.21	1.02

4. Let X be a Random variable such that $P(X=-2) = P(X=-1) = P(X=1) = P(X=2)$ and $P(X<0) = P(X=0) = P(X>0)$. Determine the Probability mass function of X, and Distribution fun. of X.

Soln:

Let, $P(X=-2) = P(X=-1) = P(X=1) = P(X=2) = a$.

$P(X<0) = P(X=-2) + P(X=-1) = a+a = 2a$.

$P(X<0) = P(X=0) = P(X>0) = 2a$.

Probability Function

$X=x$	-2	-1	0	1	2
$P(X=x)$	a	a	$2a$	a	a

i) $\sum_{i=-2}^{2} P(x_i) = 1$

$$a + a + 2a + a + a = 1$$
$$6a = 1$$
$$\boxed{a = 1/6}$$

Probability Mass function

x	-2	-1	0	1	2
$P(x)$	$1/6$	$1/6$	$2/6$	$1/6$	$1/6$

Probability Distribution:

x	$P(x)$	$F(x) = P(X \le x)$
-2	$1/6$	$1/6$
-1	$1/6$	$2/6$
0	$2/6$	$4/6$
1	$1/6$	$5/6$
2	$1/6$	1

5. If the Random variable X has the values $\{1, 2, 3, 4\}$ such that $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4)$. Find the probability distribution and cumulative distribution of X.

Soln:

Let $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4) = a$

$P(X=1) = a/2$, $P(X=2) = \frac{a}{3}$, $P(X=3) = a$, $P(X=4) = a/5$.

Probability Function:

X	1	2	3	4
$P(X)$	$a/2$	$a/3$	a	$a/5$

$$\sum_{i=1}^{4} P(x_i) = 1$$

$$\frac{a}{2} + \frac{a}{3} + a + \frac{a}{5} = 1. \Rightarrow \frac{5a}{6} + \frac{a}{5} + a.$$

$\frac{25a + 6a}{30} + a \Rightarrow \frac{31a}{30} + a \quad \frac{31a + 30a}{30} = \frac{61a}{30}$

$\frac{61a}{30} = 1 \qquad 61a = 30$

$\boxed{a = \frac{30}{61}}$

Probability mass function :

X	1	2	3	4
P(X)	15/61	10/61	30/61	6/61

Cumulative function :

x	P(x)	F(x) = P(X ≤ x)
1	15/61	15/61
2	10/61	25/61
3	30/61	55/61
4	6/61	$\frac{61}{61} = 1$.

Continuous :

8) A continuous Random Variable, has Probability Density Function,

$$f(x) = \begin{cases} k(x-1)^3 & 1 \le x \le 3 \\ 0 & o.w \end{cases}$$

since $f(x)$ is a p.d.f

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

i) Find k?

ii) Find Distribution of x.

i) $\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_1^3 k(x-1)^3\,dx = 1$.

$= k \int_1^3 (x-1)^3\,dx = 1. \Rightarrow k\left[\frac{(x-1)^3}{}\right.$

$= k\left[\frac{(x-1)^4}{4}\right]_1^3 \Rightarrow \frac{k}{4}\left[(2)^4 + 0\right] = \frac{k}{4}[16] = 1$

$\boxed{k = 1/4}$

$\therefore$ Pdf $f(x) = \begin{cases} \frac{1}{4}(x-1)^3 ; & 1 \le x \le 3 \\ 0 ; & \text{O.W} \end{cases}$

ii) <u>Cumulative Distribution:</u>

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

$$= \int_{1}^{x} \frac{1}{4}(x-1)^3\,dx$$

$$= \frac{1}{4}\left[\frac{(x-1)^4}{4}\right]_1^x \Rightarrow \frac{1}{4}\left[\frac{(x-1)^4}{4} - 0\right]$$

$$= \frac{1}{16}(x-1)^4$$

$$F(x) = \begin{cases} 0 & ; x < 1 \\ \frac{1}{16}(x-1)^4 & ; 1 \le x \le 3 \\ 1 & ; x > 3 \end{cases}$$

(7). A continuous Random variable X that can assume any value between $x=2$ & $x=5$ has a density function,

$f(x) = k(1+x)$ and find probability of $P(X<4)$.

<u>Soln:</u> Since $f(x)$ is a Pdf,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{2}^{5} k(1+x)\,dx = 1$$

$$\Rightarrow k\left[\frac{(1+x)^2}{2}\right]_2^5 = 1. \quad = \frac{k}{2}\left[\left(5+\frac{25}{2}\right) - \left(2+\frac{4}{2}\right)\right] = 1$$

$$= k\left[5+\frac{25}{2} - \cancel{2}^{4} - \cancel{2}\right] = 1$$

$$k\left[\frac{27}{2}\right] = 1 \qquad = k\left[\frac{2+25}{2}\right] = 1$$

$$\boxed{k = \frac{2}{27}}$$

$$\therefore \text{pdf } f(x) = \frac{2}{27}(1+x) ; \quad 2 < x < 5.$$

cumulative:

$$P(X<4) = \int_2^4 \frac{2}{27}(1+x)\,dx = \frac{2}{27}\left[x+\frac{x^2}{2}\right]_2^4$$

$$= \frac{2}{27}\left[(4+8)-(2+2)\right] = \boxed{\frac{16}{27} = P(X<4)}$$

(8) A continuous random variable X has density function,

$$f(x) = \frac{k}{1+x^2} \qquad -\infty < x < \infty$$

i) find the value of k

ii) find the Distribution function

soln: Since $f(x)$ is a pdf,

$$\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1 \Rightarrow \int_{-\infty}^{\infty} \frac{k}{1+x^2}\,dx = 1.$$

$$\Rightarrow k\left[\tan^{-1}x\right]_{-\infty}^{\infty} = 1.$$

$\tan^{-1}\infty = \frac{\pi}{2}$

$$\Rightarrow k\left[\tan^{-1}(\infty) - \tan^{-1}(-\infty)\right]$$

$\tan(-\theta) = -\tan\theta$

$-\tan(\infty)$

$-\frac{\pi}{2}$

$$\Rightarrow k\left[\frac{\pi}{2}+\frac{\pi}{2}\right] = 1 \Rightarrow k\pi = 1$$

$$\boxed{k = 1/\pi}$$

$\therefore$ pdf $f(x) = \dfrac{1}{\pi(1+x^2)}$; $-\infty < x < \infty$.

(ii) P. Distribution function / cumulative : D.F

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

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

$$= \frac{1}{\pi}\left[\tan^{-1}x\right]_{-\infty}^{x}$$

$$= \frac{1}{\pi}\left[\tan^{-1}(x) - \tan^{-1}(-\infty)\right]$$

$$= \frac{1}{\pi}\left[\tan^{-1}(x) + \frac{\pi}{2}\right]$$

(9). A continuous Random Variable X, has the Distribution function $F(x)$,

$$F(x) = \begin{cases} 0 & x \le 1 \\ K(x-1)^4 & 1 \le x \le 3 \\ 1 & x > 3 \end{cases}$$

i) Find K, ii) Find pdf $f(x)$
iii) Find $P(X < 2)$.

W.K.T, pdf $f(x) = \frac{d}{dx}[F(x)]$

$f(x) = 4K(x-1)^3 \; ; \; 1 \le x \le 3$

ii) Since $f(x)$ is pdf,

$\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1. \Rightarrow \int_1^3 4K(x-1)^3\,dx = 1$

$= K \int_1^3 4(x-1)^3\,dx = 1 \Rightarrow K\left[\frac{4(x-1)^4}{4}\right]_1^3 = 1.$

$= \frac{K}{\cancel{4}}\left[\cancel{4}(16)\right]$ $\quad 16K = 1$ $\quad \boxed{K = 1/16}$

since pdf

$f(x) = 4\left(\frac{1}{16}\right)(x-1)^3; 1 \le x \le 3$

$\boxed{f(x) = \frac{1}{4}(x-1)^3; 1 \le x \le 3}$

iii) pdf, $F(x)$

$F(x)$

$P(X<2) = \int_1^2 f(x)\,dx$

note: $P(X \le 2)$ & $P(X<2)$ Same

$= \int_1^2 \frac{1}{4}(x-1)^3\,dx$

$= \frac{1}{4}\left[\frac{(x-1)^4}{4}\right]_1^2 \Rightarrow \frac{1}{16}\left[1-0\right]$

$\boxed{P(X<2) = \frac{1}{16}}$

(10). If the cumulative Distribution function of Random Variable X is given by,

$$F(x) = \begin{cases} 1 - \frac{4}{x^2} & ; x > 2 \\ 0 & ; x \le 2 \end{cases}$$

i) $P(X<3)$ ii) $P(4<X<5)$ iii) $P(X \ge 3)$.

Soln: $P(X<3) = F(3) = 1 - \frac{4}{3^2} = 1 - \frac{4}{9} = \frac{5}{9}$.

$P(4<X<5) = F(5) - F(4)$

$= 1 - \frac{4}{25} - \left(1 - \frac{4}{16}\right)$

$= \frac{21}{25} - \left(\frac{12}{16}\right) \Rightarrow \frac{21}{25} - \frac{3}{4} = \frac{84-75}{100}$

$\boxed{P(4<X<5) = \frac{9}{100}}$

iii) $P(X \geq 3) = 1 - P(X<3)$

$= 1 - \frac{5}{9} = \boxed{\frac{4}{9} = P(X \geq 3)}$

(11). Let X be a continuous Random variable. its pdf

$$f(x) = \begin{cases} ax & 0 \leq x \leq 1 \\ a & 1 \leq x \leq 2 \\ -ax+3a & 2 \leq x \leq 3 \\ 0 & x>3 \end{cases}$$

i) Determine the constant 'a'.

ii) compute ~~Pro~~ $P(X \leq 1.5)$.

iii) find the cumulative D.F of X.

Soln:

$$\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (-ax+3a)\,dx + \int_3^{\infty} 0\,dx = 1$$

$$= a\left[\frac{x^2}{2}\right]_0^1 + a\left[x\right]_1^2 + a\left[-\frac{x^2}{2}+3x\right]_2^3 = 1.$$

$$= a\left[\frac{1}{2}\right] + a[1] + a\left[\left(-\frac{9}{2}+9\right) - \left(-\frac{4}{2}+6\right)\right] = 1.$$

$$= \frac{a}{2} + a + \frac{9a}{2} - \frac{8a}{2} = 1.$$

$$= \frac{3a}{2} + \frac{9a}{2} - \frac{8a}{2} = \frac{12a}{2} - \frac{8a}{2} = \frac{4a}{2} = 1$$

$\boxed{a = 1/2}$

$$\therefore \text{Pdf } f(x) = \begin{cases} \frac{x}{2} & 0 \leq x \leq 1 \\ 1/2 & 1 \leq x \leq 2 \\ -\frac{x}{2}+\frac{3}{2} & 2 \leq x \leq 3 \\ 0 & x>3 \end{cases}$$

ii) $P(X \le 1.5)$

$$\Rightarrow \int_0^{1.5} f(x)\,dx \Rightarrow \int_0^{1} \frac{x}{2}\,dx + \int_1^{1.5} \frac{1}{2}\,dx.$$

$$= \left.\frac{x^2}{4}\right]_0^1 + \frac{1}{2}\left[x\right]_1^{1.5} \Rightarrow \frac{1}{4} + \frac{1}{2}[1.5-1]$$

$$= \frac{1}{4} + \frac{1}{2}[0.5]$$

$$= \frac{1}{4} + 0.5 \Rightarrow \frac{2}{4} = \frac{1}{2}$$

$$\boxed{P(X \le 1.5) = \frac{1}{2}}$$

iii) Cumulative Distribution :

In $0 \le x \le 1$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_0^x \frac{x}{2}\,dx = \left.\frac{x^2}{4}\right]_0^x$$

$$= \frac{x^2}{4}$$

In $1 \le x \le 2$: $\int_0^1 \left(\frac{x}{2}\right)dx + \int_1^x \left(\frac{1}{2}\right)dx = \frac{x}{2} - \frac{1}{4}$.

$$F(x) = \frac{x^2}{4} + \int_1^x f(x)\,dx = \int_1^x \frac{1}{2}\,dx = \frac{x}{2} + \frac{x^2}{4}.$$

In $2 \le x \le 3$:

$$F(x) = \int_0^1 \left(\frac{x}{2}\right)dx + \int_1^2 \frac{1}{2}\,dx + \int_2^x \left(-\frac{x}{2} + \frac{3}{2}\right)dx.$$

$$= \left.\frac{x^2}{4}\right]_0^1 + \left.\frac{x}{2}\right]_1^2 + \left[-\frac{x^2}{4} + \frac{3x}{2}\right]_2^x$$

$$= \frac{1}{4} + \left(1 - \frac{1}{2}\right) + \left(-\frac{x^2}{4} + \frac{3x}{2} + 1 - 3\right]$$

$$= \frac{1}{4} + \frac{1}{2} - \frac{x^2}{4} + \frac{3x}{2} - 2$$

$$= \frac{3}{4} - 2 + \frac{3x}{2} - \frac{x^2}{4}$$

$$= -\frac{5}{4} + \frac{3x}{2} - \frac{x^2}{4}$$

$$= -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4}$$

In $x > 3$:

$$F(x) = \int_0^1 \frac{x}{2}\,dx + \int_1^2 \frac{1}{2}\,dx + \int_2^3 \left(-\frac{x}{2} + \frac{3}{2}\right)dx + \int_3^x (0)\,dx$$

$$= \frac{x^2}{4}\Big]_0^1 + \frac{x}{2}\Big]_1^2 - \frac{x^2}{4} + \frac{3x}{2} - 2$$

$$= \frac{1}{4} + \frac{1}{2} - \frac{x^2}{4} + \frac{3x}{2}\Big)_2^3$$

$$= \frac{3}{4} - \frac{9}{4} + \frac{9}{2} + 4 - 3$$

$$= \frac{3}{4} - 2 + \frac{9}{4} = \frac{12}{4} - 2 = \frac{4}{4} = 1.$$

$$= 1$$

$$F(x) = \begin{cases} 0 & ; x < 0 \\ x^2/4 & ; 0 \le x \le 1 \\ \frac{x}{2} - 1/4 & ; 1 \le x \le 2 \\ -\frac{x^2}{4} + \frac{3x}{2} - \frac{5}{4} & ; 2 \le x \le 3 \\ 1 & ; x > 3 \end{cases}$$

CHENNAI INSTITUTE OF TECHNOLOGY * 69 * CHENNAI DEPT. OF COMPUTER SCIENCE AND ENGINEERING

A ①. If x is a continuous Random variable with Pdf as $f(x) = \begin{cases} x & 0 \le x \le 1 \\ 3/2\,(x-1)^2 & 1 \le x \le 2 \\ 0 & x > 2 \end{cases}$

Find, i) cumulative D.F & using to find $P\left(\frac{3}{2} < x < \frac{5}{2}\right)$

8mk 16m ②. The diameter of an Electric cable say X is assumed to be continuous Random Variable Pdf $f(x) = 6x(1-x)$; $0 \le x \le 1$. $\int_0^b 6x(1-x)\,dx = \int_b^1 6x(1-x)\,dx$.

i) check that the above is a Pdf.

ii) Determine 'b', such $P(X < b) = P(X > b)$ $2, 0.5$

iii) Find the C.D.F of X. $\int_{-\infty}^{x} \to \int_0 f(x)\,dx$

iv) Find $P\left(X \le 1/2 \,/\, 1/3 < X < 2/3\right)$

H.W

(1). $f(x) = \begin{cases} x & 0 \le x \le 1 \\ \frac{3}{2}(x-1)^2 & 1 \le x \le 2 \\ 0 & x > 2 \end{cases}$

Cumulative distribution:

In $0 \le x \le 1$ $F(x) = \int_0^x f(x)\,dx = \int_0^x x\,dx = \left.\frac{x^2}{2}\right]_0^x = \frac{x^2}{2}$

In $1 \le x \le 2$.

$$F(x) = \int_0^1 x\,dx + \int_1^x \frac{3}{2}(x-1)^2\,dx$$

$$= \left.\frac{x^2}{2}\right]_0^1 + \frac{3}{2}\left[\frac{(x-1)^3}{3}\right]_1^x$$

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

In $x > 2$:

$$F(x) = \int_0^1 x\,dx + \int_1^2 \frac{3}{2}(x-1)^2\,dx + \int_2^x 0\,dx.$$

$$= \left.\frac{x^2}{2}\right]_0^1 + \frac{3}{2}\left[\frac{(x-1)^3}{3}\right]_1^2$$

$$= \frac{1}{2} + \frac{\cancel{3}}{2}\left[\frac{1}{\cancel{3}}\right] \Rightarrow \frac{1}{2} + \frac{1}{2} = 1$$

$$F(x) = \begin{cases} x^2/2 & 0 \le x \le 1 \\ \frac{1}{2} + \frac{(x-1)^3}{2} & 1 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

$$P\left(\frac{3}{2} < x < \frac{5}{2}\right) = F\left(\frac{5}{2}\right) - F\left(\frac{3}{2}\right)$$

$$\Rightarrow \int \Rightarrow 1 - \left(\frac{1}{2} + \left(\frac{\left(\frac{3}{2}-1\right)^3}{2}\right)\right)$$

$$\Rightarrow 1 - \left(\frac{1}{2} + \frac{\left(\frac{1}{2}\right)^3}{2}\right) \Rightarrow 1 - \frac{1}{2} + \frac{1}{16}$$

$$= 1 - \frac{9}{16} = \frac{7}{16}.$$

$$\boxed{P\left(\frac{3}{2} < x < \frac{5}{2}\right) = \frac{7}{16}}$$

(2). $f(x) = 6x(1-x)$; $0 \le x \le 1$.

i) Above pdf / Not,

$\Rightarrow \int_0^1 6x(1-x)\,dx \Rightarrow \int_0^1 6x - 6x^2\,dx \Rightarrow \left[\frac{6x^2}{2} - \frac{6x^3}{3}\right]_0^1$

$= \left[\frac{6}{2} - \frac{6}{3}\right] = 3-2 = 1$ //.

The above $f(x)$ is a pdf.

ii) $b = ?$ $P(X < b) = P(X > b)$

$f(x) = 6x(1-x)$; $0 \le x \le 1$

$\Rightarrow \int_0^b f(x)\,dx = \int_b^1 f(x)\,dx$

$\Rightarrow \int_0^b \cancel{6x}(1-x)\,dx = \int_b^1 \cancel{6x}(1-x)\,dx$

$\Rightarrow \int_0^b (1-x)\,dx = \int_b^1 (1-x)\,dx$

$\Rightarrow \left[x - \frac{x^2}{2}\right]_0^b = \left[x - \frac{x^2}{2}\right]_b^1$

$\Rightarrow b - \frac{b^2}{2} = \left(1 - \frac{1}{2}\right) - \left(b - \frac{b^2}{2}\right)$

$b - \frac{b^2}{2} = \left(\frac{1}{2}\right) - b + \frac{b^2}{2}$

$b + b - \frac{b^2}{2} - \frac{b^2}{2} = \frac{1}{2} \Rightarrow 2b - \frac{\cancel{2}b^2}{\cancel{2}} = \frac{1}{2}$

$-b^2 + 2b = \frac{1}{2}$ $\qquad -b^2 + 2b - \frac{1}{2} = 0.$

$b^2 - 2b + \frac{1}{2} = 0.$

$b = \frac{1}{2}$ is only possible

iii) C.D.F $\rightarrow \int_{-\infty}^{x} f(x)\,dx \Rightarrow \int_0^x f(x)\,dx \Rightarrow \int_0^x 6x(1-x)\,dx$

$\Rightarrow \left[\frac{6x^2}{2} - \frac{6x^3}{3}\right]_0^x \Rightarrow \boxed{3x^2 - 2x^3}$.

iv) $P\left(X \le \frac{1}{2} \Big/ \frac{1}{3} < X < \frac{2}{3}\right) = P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$

Mathematical Expectations and Moments:

Expectations:

The average process when applied to the Random Variable is called Expectation. It is denoted by $E[X]$.

(Mean) $E[X] = \sum x_i \, P(x_i)$ [Discrete case].

$E[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx$ [continuous case].

Properties: Expectations:

i) $E(a) = a$; a is constant.

ii) $E[ax+b] = aE[X] + b$.

iii) $E[aX \pm bY] = aE[X] \pm bE[Y]$

iv) $E[XY] = E[X]\,E[Y]$. (X, Y indepeut)

Note:

Mean of $X = \bar{X} = E[X] = \mu_1'$

Variance of $X = E[X^2] - [E(X)]^2 \Rightarrow \boxed{\mu_2 = \mu_2' - \mu_1'^2}$

Properties of Variance

2m

i) $\text{Var}\, a = 0$ a is constant.

ii) $\text{Var}(ax+b) = a^2 \,\text{Var}\, X$

iii) $\text{Var}(aX \pm bY) = a^2 \,\text{Var}\, X \pm b^2 \,\text{Var}\, Y$.

Moments:

The r^{th} moment about the origin of Random Variable X, defined as, expected Value r^{th} power of X.

$r=1$ Mean

$r=2$ Mean square

The r^{th} moment about the origin

$$E[X^r] = \sum x_i^r p(x_i) \quad \text{(Discrete case)}$$

$$E[X^r] = \int_{-\infty}^{\infty} x^r f(x)\,dx \quad \text{(continuous case)}.$$

- The r^{th} moment about the point,

$$E[(X-A)^r] = \sum (x_i - A)^r p(x_i) \quad [\text{Discrete case}]$$

$$E[(X-A)^r] = \int_{-\infty}^{\infty} (x-A)^r f(x)\,dx \quad [\text{continuous case}].$$

The r^{th} moment about the Mean,

$$E[(X-\bar{X})^r] = \sum (x_i - \bar{x})^r p(x_i) \quad [\text{Discrete case}]$$

$$E[(X-\bar{X})^r] = \int_{-\infty}^{\infty} (x-\bar{x})^r f(x)\,dx \quad [\text{continuous case}]$$

Problems:

(1) Let X be a Random Variable with, $E[X] = 1$, $E[X(X-1)] = 4$. Find Var X, Var(2-3X), Var$\left(\frac{X}{2}\right)$.

Given: $E[X] = 1$,
$$E[X(X-1)] = 4$$
$$E[X^2 - X] = 4$$
$$E[X^2] - E[X] = 4.$$
$$E[X^2] = 4 + E[X]$$
$$E[X^2] = 4+1 = \boxed{5 = E[X^2]}$$

$$\text{Var of } X = E[X^2] - [E(X)]^2$$
$$= 5 - 1 = 4$$

ii) $\text{Var}[2-3X] = (-3)^2 \text{ Var } X = 9 \times 4 = 36$,

iii) $\text{Var}\left(\frac{X}{2}\right) = \left(\frac{1}{2}\right)^2 \text{ Var } X = \frac{1}{\cancel{4}} \times \cancel{4} = 1.$

(8m)

(2). The cumulative distribution of Random Variable X is, $F(x) = 1-(1+x)e^{-x}$, $x>0$ $0<x<\infty$. Find the pdf of X, Mean, and Var of X.

$$F(x) = 1-(1+x)e^{-x}$$
$$= 1 - e^{-x} - xe^{-x}$$

The pdf $f(x) = \frac{d}{dx}[F(x)]$

$$= 0 + e^{-x} - [-xe^{-x} + e^{-x}(1)]$$
$$= 0 + \cancel{e^{-x}} + xe^{-x} - \cancel{e^{-x}}$$
$$f(x) = xe^{-x}$$

$e^{-\infty} = 0$

$e^{0} = 1$

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

$$= \int_{0}^{\infty} x.(xe^{-x})\,dx \Rightarrow \int_{0}^{\infty} x^2 e^{-x}\,dx$$

$u = x^2$, $v = e^{-x}$

$u' = 2x$, $v_1 = e^{-x}/-1$

$u'' = 2$, $v_2 = \frac{e^{-x}}{1}$

$v_3 = \frac{-e^{-x}}{1}$

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

$$= [0-(-2)] = \boxed{2 = \text{Mean}} = E[X]$$

~~Var of x~~ $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$

$$= \int_0^{\infty} x^2 . xe^{-x} dx \Rightarrow \int_0^{\infty} x^3 e^{-x} dx$$

$u = x^3$, $v = e^{-x}$

$u' = 3x^2$, $v_1 = -e^{-x}$

$u'' = 6x$, $v_2 = e^{-x}$

$u''' = 6$, $v_3 = -e^{-x}$

$v_4 = e^{-x}$

$$= \left[x^3(-e^{-x}) - 3x^2(e^{-x}) - 6x(e^{-x}) - 6e^{-x}\right]_0^{\infty}$$

$$= [0-(-6)] = 6$$

Var of $X = E[X^2] - E[X]^2$

$$= 6 - (2)^2 = 6-4 = \boxed{2 = \text{Var of } x}$$

(8m)

(3) A continuous Random variable X has pdf $f(x) = Kx^2e^{-x}$ $x>0$, Find K of r^{th} moment, mean and variance.

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

$\Rightarrow \int_0^{\infty} K x^2 e^{-x}\,dx = 1 \Rightarrow K \int_0^{\infty} x^2 e^{-x}\,dx = 1.$

$\Rightarrow K\left[x^2(-e^{-x}) - 2x(e^{-x}) + 2(-e^{-x})\right]_0^{\infty} = 1.$

$= K[0-(-2)] = 1 \qquad 2K = 1 \qquad \boxed{K = \frac{1}{2}}$ $\boxed{\text{Pdf } f(x) = \frac{1}{2}x^2 e^{-x}, x>0}$

$E[X^r] = \int_{-\infty}^{\infty} x^r f(x)\,dx = $ rth moment

$= \int_0^{\infty} x^r \cdot \frac{1}{2} x^2 e^{-x}\,dx$

$\boxed{\text{Gamma fun. } \Gamma n = \int_0^{\infty} e^{-x} x^{n-1}\,dx}$

$= \frac{1}{2}\int_0^{\infty} x^{(r+2)} e^{-x}\,dx = \frac{1}{2}\int_0^{\infty} x^{(r+3)-1} e^{-x}\cdot dx$

$\Gamma n = (n-1)!$

$= \frac{1}{2}\Gamma r+3 \quad = \boxed{\frac{1}{2}(r+2)! = E[X^r]}$

Mean: $E[X] = \int_{-\infty} \frac{1}{2}(3)! = \frac{6}{2} = \boxed{3 = \text{Mean}}$
$r=1$

Variance: $E[X^2] = \frac{1}{2}(4)! = \frac{24}{2} = 12$
$r=2$

$\text{Var of } X = E[X^2] - (E[X])^2$

$= 12 - 9 = \boxed{3 = \text{Var of } X}$

2m

4. If X and Y are independent Random variable with Variance 2 & 3, Find Var (3X+4Y).

$E[X^2] - [E(X)]^2$.

Soln:

Given : $Var(X) = 2$, $Var(Y) = 3$.

$Var(3X+4Y) = 3^2 Var(X) + 4^2 Var(Y)$

$= 9(2) + 16(3)$

$= 18 + 48$

$= 66$ //.

Moment Generating function (M.g.f) :- $M_X(t)$.

It is used to calculate higher Moments

$$M_x(t) = E[e^{tx}] = \sum e^{tx} p(x_i) \quad [D.C]$$

$$M_x(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad [C.C]$$

Note:-

$$E[e^{tx}] = E\left[1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \dots + \frac{(tx)^r}{r!} + \dots\right]$$

$$= 1 + E(x)\left(\frac{t}{1!}\right) + E(x^2)\left(\frac{t^2}{2!}\right) + \dots + E(x^r)\left(\frac{t^r}{r!}\right) + \dots$$

$$= 1 + \mu_1'\left(\frac{t}{1!}\right) + \mu_2'\left(\frac{t^2}{2!}\right) + \dots + \mu_r'\left(\frac{t^r}{r!}\right) + \dots$$

μ_r' = r^{th} moment about the origin

the co-efficient of $\frac{t^r}{r!}$ it gives μ_r'.

Properties of MGF:

i) $E(X) = [M_X'(t)]_{t=0} = \left[\frac{d}{dt} M_X(t)\right] = M_X'(t)_{t=0}$.

ii) $E(X^2) = [M_X''(t)]_{t=0}$.

1. Find the M.g.f of Random Variable whose Probability function $P(X=x) = \frac{1}{2^x}$; $x = 1, 2, 3, \dots$ Hence find its Mean.

Hint - discrete - It is Countable, it is Countably infinite.

Sol:

$P(X=x) = \frac{1}{2^x}$.

$$Mgf = M_x(t) = E[e^{tx}] = \sum_{i=0}^{\infty} e^{tx} p(x_i)$$

$$= \sum_{i=1}^{\infty} e^{tx} \left(\frac{1}{2^x}\right)$$

$$= \sum_{i=1}^{\infty} \left(\frac{e^t}{2}\right)^x$$

$$= \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \cdots$$

$$= \frac{e^t}{2}\left[1 + \left(\frac{e^t}{2}\right)^1 + \left(\frac{e^t}{2}\right)^2 + \cdots\right]$$

$$= \frac{e^t}{2}\left[1 - \frac{e^t}{2}\right]^{-1}$$

$$= \frac{e^t}{2}\left[\frac{2-e^t}{2}\right]^{-1}$$

$$(1-x)^{-1} = \frac{1}{1-x}.$$

$$= \frac{e^t}{\cancel{2}}\left[\frac{\cancel{2}}{2-e^t}\right]$$

$$M_x(t) = \frac{e^t}{2-e^t} \qquad \frac{u}{v} \quad \frac{vu'-uv'}{v^2}$$

$$M'_x(t) = \frac{(2-e^t)(e^t) - e^t(-e^t)}{(2-e^t)^2}$$

$$= \frac{2e^t - e^{2t} + e^{2t}}{(2-e^t)^2}$$

$$= \frac{2e^t}{(2-e^t)^2}$$

<u>Mean:</u>

$$E(X) = M'_x(0) = \frac{2}{(2-1)^2} = \frac{2}{1} = 2$$

② If a Random variable X has m.g.f $M_x(t) = \frac{2}{2-t}$ Find the Var x.

$$M_x(t) = \frac{2}{2-t} = 2(2-t)^{-1}$$

$$M'_x(t) = -2(2-t)^{-2}(-1) = 2(2-t)^{-2}$$

$$M''_x(t) = -4(2-t)^{-3}(-1) = 4(2-t)^{-3}.$$

$$E(X) = M'_x(0) = 2(2-0)^{-2} = \frac{2}{2^2} = \frac{1}{2}.$$

$$E(X^2) = M''_x(0) = 4(2-0)^{-3} = \frac{4}{2^3} = \frac{1}{2}$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

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

$$= \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

Q. If x has the distribution function,

$$F(x) = \begin{cases} 0 & x \le 1 \\ 1/3 & 1 \le x \le 4 \\ 1/2 & 4 \le x \le 6 \\ 5/6 & 6 \le x \le 10 \\ 1 & x \ge 10 \end{cases}$$

i) Find p.d.f $f(x)$
ii) Find $P(2 < x < 6)$
iii) Find mean of x
iv) Find variance of x.

Given problem is discrete and can't satisfy $f(x) = F'(x)$.

Probability Distribution:

x	P(X)	$F(X) = P(X \le x)$
0	0	0
1	1/3	1/3
4	1/6	1/2
6	2/6	5/6
10	1/6	1.

$X=x$	0	1	4	6	10
$P(X=x)$	0	1/3	1/6	2/6	1/6.

i) $P(2 < x < 6) = P(X=4) = 1/6$.

ii) Mean of x $= E[x] = \sum x_i p(x_i) = (0 \times 0) + (1 \cdot \frac{1}{3}) + (4 \times \frac{1}{6}) + (6 \times \frac{5}{6}) + (10 \times \frac{1}{6})$

$$= \frac{1}{3} + \frac{2}{3} + 2 + \frac{5}{3} \qquad \boxed{E[X] = \frac{14}{3}}$$

$E[x^2] = \sum x_i^2 p(x_i) = (0.0) + (1 \times \frac{1}{3}) + (16 \times \frac{1}{6}) + (36 \times \frac{2}{6}) + (100 \times \frac{1}{6})$

$$= \frac{1}{3} + \frac{8}{3} + 12 + \frac{50}{3}$$

$$= \frac{1+8+36+50}{3} = \frac{95}{3}.$$

$Var[x] = E[x^2] - [E(x)]^2 = \frac{95}{3} - \frac{196}{9} = \frac{285-196}{9} = \frac{89}{9}.$

$$\boxed{Var[x] = \frac{89}{9}.}$$

Q. If x has a Random Variable,

$$f(x) = \begin{cases} \frac{1}{3} e^{-x/3}, & x > 0 \\ 0 & \text{o.w.} \end{cases}$$

find
i) $P(X>3)$
ii) mgf
iii) $E(X)$
iv) $Var(x)$.

Soln:

$$mgf = M_x(t) = E[e^{tx}] = \int_0^\infty e^{tx} \cdot \frac{1}{3} e^{-x/3}\, dx.$$

$$= \frac{1}{3} \int_0^\infty e^{tx - x/3}\, dx \Rightarrow \frac{1}{3} \int_0^\infty e^{-x\left(\frac{1}{3} - t\right)} dt.$$

$$= \frac{1}{3} \left[\frac{e^{-x\left(\frac{1}{3} - t\right)}}{-\left(\frac{1}{3} - t\right)} \right]_0^\infty$$

$$= \frac{1}{\cancel{3}} \left[0 + \frac{1}{\frac{1-3t}{\cancel{3}}} \right] = \frac{1}{1-3t} = (1-3t)^{-1}.$$

$$M_x(t) = (1-3t)^{-1}$$

$$M_x'(t) = -1(1-3t)^{-2}(-3) = 3(1-3t)^{-2}.$$

$$M_x''(t) = -6(1-3t)^{-3}(-3) = 18(1-3t)^{-3}.$$

$$E(X) = M_x'(0) = 3(1-0)^{-2} = 3$$

$$E(X^2) = M_x''(0) = 18(1-0)^{-3} = 18.$$

$$Var\ x = E[X^2] - (E[X])^2 = 18 - (3)^2 = 18 - 9 = 9.$$

$$\boxed{Var\ x = 9.}$$

(5) If a Random variable, X has mgf,

$M_x(t) = \frac{2}{2-t}$, Determine variance of x,

Soln: $M_x(t) = \frac{2}{2-t} = 2(2-t)^{-1}$

$M_x'(t) = -2(2-t)^{-2}(-1) = 2(2-t)^{-2} = 2(2)^{-2} = \frac{2}{(2)^2} = \frac{1}{2}$

$M_x''(t) = -4(2-t)^{-3}(-1) = 4(2-t)^{-3} = 4(2)^{-3} = \frac{4}{(2)^3} = \frac{4}{8}$

$E(x) = M_x'(0) = \cancel{4}\ \frac{1}{2}$

$E(x^2) = M_x''(0) = \cancel{8}\ \frac{1}{2}$.

$Var\ x = E(x^2) - [E(x)]^2 = \cancel{8 - (4)^2 = 8 - 16}$

$= \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \boxed{\frac{1}{4}}$ $= \cancel{-8}$.

(8). PROBABILITY DISTRIBUTIONS :

1. Binomial distribution
2. Poisson distribution
3. Geometric distribution.

} Discrete (p.m.f)

4. Uniform distribution
5. Exponential distribution
6. Gamma distribution
7. Weibull distribution
8. Normal distribution.

} continuous. (p.d.f) density.

Binomial Distribution :

A Random Variable x is said to be, follow, Binomial distribution, if it only assume non-negative values with Probability Mass function is given by,

p.m.f $P(X=x) = nC_x\, p^x q^{n-x}$, $x = 0, 1, 2, \ldots n$.

where n = number of trials

x = no. of success

$n-x$ = no. of failures.

Note:

$q + p = 1$.

The parameters are, $B(n, p)$

Binomial Frequency distribution.

$P(X=x) = N.\ nC_x p^x q^{n-x}$; $x = 0, 1, 2, 3 \ldots n$.

Properties;

i) Each Trials are Bernoulli's (either success/ Failure)

ii) The no. of trails 'n' is finite (or) is a small value.

iii) The trails are independent of each other.

iv) The probability of success 'p' is constant for each trail. [fixed].

(S). Find Mean, Variance, and Mgf of Binomial distribution.

soln:

The probability mass function of BD. is,

$P(X=x) = nC_x \cdot p^x q^{n-x}$; $x = 0, 1, 2, 3 \ldots n$.

The mgf $= M_X(t) = E[e^{tx}] = \sum_{x=0}^{n} e^{tx} p(x)$

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

$= \sum_{x=0}^{n} nC_x\, (pe^t)^x q^{n-x}$.

$= (q+pe^t)^n$ [By Binomial Th].

Binomial Th,

$(x+a)^n = nC_0 x^n a^0 + nC_1 x^{n-1} a^1 + \ldots nC_n x^0 a^n$

$= \sum_{r=0}^{n} nC_r x^r a^{n-r}$.

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

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

$M_X'(t) = $ Mean $= n(q+pe^t)^{n-1} \cdot pe^t$.

$M_X''(t) = np[(q+pe^t)^{n-1} \cdot e^t + e^t (n-1)(q+pe^t)^{n-2} pe^t]$

$E[X] = M_X'(0) = n(q+q)^{n-1} p = n(1)p = \boxed{np = E[X]}$

$\therefore \boxed{q+p=1}$.

$E[X^2) = M_X''(0) = np[(q+p)^{n-1} + (n-1)(q+p)^{n-2} p]$

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

$= np[1+np-p]$ $\boxed{1-p=q}$

$= np[q+np]$

$E[X^2] = npq + n^2p^2$.

$\boxed{Var(X) = E[X^2] - (E[X])^2}$

$= npq + n^2p^2 - (np)^2$

$\boxed{Var(X) = npq}$.

(2). (4)(5). The Mean of a Binomial Distribution is 20, and its S.D is '4'. Determine the parameters of the Distribution.

Soln:

$$\text{Mean} = E(x) = np = 20 \quad - (1)$$

$$SD = \sqrt{\text{var of}(x)} = \sqrt{npq} = 4$$

$$npq = 16 \quad - (2).$$

$$\frac{(2)}{(1)} = \frac{npq}{np} = \frac{16}{20}$$

$$\boxed{\begin{matrix} p+q=1 \\ p=1-q \end{matrix}}$$

$$q = \frac{4}{5}$$

$$p = 1-\frac{4}{5} \Rightarrow \boxed{p=\frac{1}{5}}$$

$$n\left(\frac{1}{5}\right) = 20 \quad \boxed{n=100}$$

The p.m.f of B.D is,

$$\boxed{P(X=x) = 100C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{100-x}}.$$

(8) For the triangular distribution, $f(x) = \begin{cases} x & 0 < x \le 1 \\ 2-x & 1 \le x \le 2 \\ 0 & \text{o.w.} \end{cases}$

Find Mean, Variance, Mgf.

$$\text{Mean} = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x^2\,dx + \int_1^2 x(2-x)\,dx$$

$$= \left[\frac{x^3}{3}\right]_0^1 + \int_1^2 (2x - x^2)\,dx \Rightarrow \frac{1}{3} + \left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_1^2$$

$$= \frac{1}{3} + \left[4 - \frac{8}{3} - 1 + \frac{1}{3}\right]$$

$$= \frac{2}{3} - \frac{8}{3} + 3 \Rightarrow -\frac{6}{3} + 3 = -2+3 = 1$$

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

$$= \int_0^1 x^3\,dx + \int_1^2 x^2(2-x)\,dx \Rightarrow \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2$$

$$= \left[\frac{1}{4} + \frac{16}{3} - \frac{16}{4} - \frac{2}{3} + \frac{1}{4}\right]$$

$$= \frac{2}{4} - \frac{16}{4} + \frac{14}{3} = -\frac{14}{4} + \frac{14}{3} = \frac{-42+56}{12}$$

$$= \frac{14}{12} = \boxed{\frac{7}{6}}.$$

$$\underline{\text{mgf}}: \quad M_x(t) = E[e^{tx}] = \left[\int_0^1 x e^{tx} + \int_1^2 (2-x) e^{tx}\right]$$

$$\Rightarrow \left[x\left(\frac{e^{tx}}{t}\right) - \frac{e^{tx}}{t^2}\right]_0^1 + \left[(2-x)\frac{e^{tx}}{t} + \frac{e^{tx}}{t^2}\right]_1^2$$

$$\Rightarrow \left[\cancel{\frac{e^t}{t}} - \frac{e^t}{t^2} + \frac{1}{t^2} + \frac{e^{2t}}{t^2} - \cancel{\frac{e^t}{t}} - \frac{e^t}{t^2}\right]$$

$$\Rightarrow \frac{1 + e^{2t} - 2e^t}{t^2}$$

$$\Rightarrow \frac{(1-e^t)^2}{t^2} \qquad \therefore \frac{(1-e^t)^2}{t^2}$$

(2). The Mean of a Binomial Distribution is 20, and its S.D is '4'. Determine the parameters of the Distribution.

(Q) A Discrete Random Variable, x has, Mgf $M_x(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$ find $E(x)$, $Var(x)$, $P(X = 2)$.

$\underline{\text{Soln:}}$ In Binomial Distribution,

$$\text{mgf}: \boxed{M_x(t) = (q + pe^t)^n} = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$$

$$p = \frac{3}{4}, \quad q = 1/4, \quad n = 5$$

$\therefore$ The Probability Mass func.

$$P(X = x) = 5C_x \left(\frac{3}{4}\right)^x \left(\frac{1}{4}\right)^{5-x}, \quad x = 0, 1, 2, \ldots 5.$$

i) $E(x) = np = 5\left(\frac{3}{4}\right) = \frac{15}{4}$.

$$\text{Var of } (x) = npq = 5\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = \frac{15}{16}.$$

$$P(X = 2) = 5C_2 \left(\frac{3}{2}\right)^2 \left(\frac{1}{4}\right)^3.$$

$5C_2$. cal 5 shift nCR $\downarrow$ C_2

$$= 5C_2 \left(\frac{9}{4}\right)\left(\frac{1}{64}\right)$$

$$= 10 \times \frac{9}{4} \times \frac{1}{64} = 0.087.$$

(2). If 10% of screws produced by automatic machines, are defective, find the probability that 20 screws selected at random,

i) exactly 2 defective
ii) Atmost 3 defective.
iii) Atleast 2 defective
iv) Between 1 and 3 defective (inclusive).

Soln, In Binomial distribution,

Giv, p(screws are defective) $= p = 10\% = \frac{10}{100} = 0.1$.

$q = 1 - 0.1 = 0.9$.

$n = 20$,

$\therefore$ The p.m.f $P(X=x) = 20C_x (0.1)^x (0.9)^{20-x}$ $x = 0 \ldots 20$.

i) $P(\text{exactly 2 defective}) = P(X=2)$

$= 20C_2 (0.1)^2 (0.9)^{18} = 0.285$

ii) $P(\text{at most 3 defective}) = P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 20C_0 (0.1)^0 (0.9)^{20} + 20C_1 (0.1)^1 (0.9)^{19} + 20C_2 (0.1)^2 (0.9)^{18} + 20C_3 (0.1)^3 (0.9)^{17}$.

$= 0.121 + 0.270 + 0.285 + 0.190$.

$= 0.86$.

iii) $P(\text{at least 2 defective}) = P(X \ge 2) = 1 - P(X < 2)$

$= 1 - [P(X=0) + P(X=1)]$

$= 1 - [0.121 + 0.270]$

$= 0.609$

iv) $P(\text{Between 1 and 3 defective}) = P(1 \le X \le 3)$

$= [P(X=1) + P(X=2) + P(X=3)]$

$= 0.270 + 0.285 + 0.190$.

$= 0.745$.

(3). A mean and variance of Binomial variate 8 & 6 respectively. Find $P(X \ge 2)$

$np = 8$ — (1)

$npq = 6$. — (2)

$\frac{(2)}{(1)} = \frac{npq}{np} = \frac{6}{8}$ $\boxed{q = \frac{3}{4}}$

$\boxed{p = \frac{1}{4}}$

p.m.f ⇒ $n\left(\frac{1}{4}\right) = 8$ $\boxed{n = 32}$.

p.m.f: $32C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{32-x} = P(X=x)$ m< l>1

$P(X \geq 2) = 1 - P(X<2)$
$= 1 - [P(X=0) + P(X=1)]$
$= 1 - \left[32C_0 \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^{32} + 32C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{31}\right]$
$= 1 - [0.0001 + 0.0011]$
$= 1 - [0.0012] = 0.9988$

(6) 6 Dice are thrown 729 times, how many times do you expect, atleast 3 dice to show a 5 or 6.

$P(X \leq 2)$.

Soln: In Binomial Distribution,

$P(\text{getting } 5 \text{ (or) } 6) = P = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$ $P(X\geq$

$q = 1 - \frac{1}{3} = \frac{2}{3}$ $\boxed{n=6}$ atleast ≥ no ≤

P.m.f: $P(X=x) = 6C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}$ $x = 0,1,2,..6$

$P(\text{atleast } 3 \text{ dice}) = P(X \geq 3) = 1 - P(X<3)$.
$= 1 - [P(X=0) + P(X=1) + P(X=2)]$
$= 1 - \left[6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 + 6C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 + 6C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4\right]$

$= 1 - [0.088 + 0.263 + 0.329]$
$= 0.32$.

6 Dice are, thrown, 729 times,

$\therefore$ pmf $= N \times P(X=x)$
$= 729 \times 0.32$
$= 233.28$.

Expect no. of times at 3 dice to show 5 (or) 6. is, 233 times

Poisson distribution:

2M Df. a distribution

If x is a Discrete Random Variable, that assumes the values, 0, 1, 2 ....etc such that the probability mass function is given by,

$$\boxed{P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} ; x = 0, 1, 2 \ldots \infty}$$

Properties:

* The No. of trials is infinitely large. $n \to \infty$.
* The Probability of success is each trial is very small. $p \to 0$.
* $np = \lambda$, λ is a parameter.

①. Find mean, Variance and mgf of Poisson distribution.

Soln: The p.m.f of Poisson distribution

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \ldots$$

D → mgf
I → M, Var

$$Mgf = M_X(t) = E[e^{tx}] = \sum_{x=0}^{\infty} e^{tx} p(x_i).$$

$$= \sum_{x=0}^{\infty} e^{tx} \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

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

$$\boxed{e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots}$$

$$= e^{-\lambda}\left[1 + \left(\frac{\lambda e^t}{1!}\right) + \left(\frac{\lambda e^t}{2!}\right)^2 + \ldots\right]$$

$$= e^{-\lambda} \cdot e^{\lambda e^t} = e^{-\lambda + \lambda e^t} = e^{-\lambda[1-e^t]} = \boxed{e^{\lambda[e^t - 1]}}$$

$$\boxed{Mgf = e^{-\lambda[1-e^t]}}$$

$$M_X'(t) = \cancel{e^{-\lambda}} [\cancel{\lambda} - e^{-\lambda(1-e^t)} (+\lambda e^t)].$$

$$M_X'(t) = \underbrace{e^{\lambda[e^t-1]}}_{u} \cdot \underbrace{\lambda e^t}_{v}$$

$$M_X''(t) = \lambda\left[e^{\lambda(e^t-1)} e^t + e^t e^{\lambda(e^t-1)} \lambda e^t\right]$$

$E[X] = M_x'(0) = \boxed{\lambda = \text{Mean}}$

$E[X^2] = M_x''(0) = \lambda(1+\lambda) = \lambda+\lambda^2$. N is large

$\text{Var of } (x) = E[x^2] - [E(x)]^2 = \lambda+\lambda^2-\lambda^2$

$= \boxed{\lambda = \text{variance}}$

(2). If 3% of electric bulbs manufactured by the company are defective find P (that 100 bulbs selected at random exactly 5 bulbs are defective).

= In poisson distribution,

P(bulbs are defective) = $p = 3\% = \frac{3}{100} = 0.03$

where $n=100$, $\lambda = np = 100 \times 0.03 = 3$.

The p.m.f of P.D is $P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$ $x = 0,1,2 \ldots$

$P(X=x) = \frac{e^{-3} 3^x}{x!}$ $x = 0,1,2 \ldots$

P(exactly 5 bulbs defective) $= \frac{e^{-3} 3^5}{5!} = 0.101$.

$\boxed{\text{P(exactly 5 bulbs defective)} = 0.101}$.

(4) The atoms of the Radio active element randomly disintergrating. if every gram this element ~~of~~ Average emits, 3.9 α particles/sec, what is the Probability that during the Next second no, of particles emitted from 1g is,

i) At most 6. ii) At least 2 iii) Atleast 3 and Atmost 6.

Soln In P.D given. λ = Average = Mean = 3.9

The p.m.f $P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-3.9}(3.9)^x}{x!}, x = 0,1,2,3$

i) P (atmost 6) $= P(X \le 6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6)$

$$\Rightarrow e^{-3.9}\left[1+\frac{3.9}{1!}+\frac{(3.9)^2}{2!}+\frac{(3.9)^3}{3!}+\frac{(3.9)^4}{4!}+\frac{(3.9)^5}{5!}+\frac{(3.9)^6}{6!}\right.$$

$$\Rightarrow e^{-3.9}\left[44.426\right] \Rightarrow 0.899.$$

$$\boxed{P(\text{atmost } 6) = 0.899}$$

ii) $P(\text{Atleast } 2) = P(X \geq 2) = 1 - P(X < 2)$

$$= 1 - [P(X=0) + P(X=1)]$$

$$= 1 - \left[e^{-3.9}\frac{(3.9)^0}{0!} + e^{-3.9}\frac{(3.9)^1}{1!}\right]$$

$$= 1 - [0.09]$$

$$\boxed{P(\text{Atleast } 2) = 0.901}$$

iii) $P(\text{atleast } 3 \text{ \& atmost } 6) = P(3 \leq X \leq 6)$.

$$= P(X=3) + P(X=4) + P(X=5) + P(X=5).$$

$$= e^{-3.9}[9.885 + 9.639 + 7.518 + 4.88]$$

$$\boxed{P(3 \leq X \leq 6) = 0.646}$$

③.ⓑ In a company of Monthly breakdown of a Machine is Random Variable with Poisson, Average 1.8, Find the probability that Machine will be function for a Month.

i) Without breakdown ii) With exactly 1 breakdown
iii) With atleast 1 breakdown.

Soln:

Average = Mean = λ = 1.8.

The P.mf, $P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1.8}(1.8)^x}{x!}$, $x = 0, 1, 2 \ldots$

i) P(without BD) $P(X=0) = \frac{e^{-1.8}(1.8)^0}{0} = 0.165$.

ii) P(with exactly 1) $P(X=1) = \frac{e^{-1.8}(1.8)^1}{1!} = 0.165 \times$
$= 0.297$.

iii) P(atleast 1 breakdown) = $P(X \geq 1) = 1 - P(X < 1)$

$= 1 - 0.297 \cdot 1 - [P(X=0)]$ P

$= 1 - 0.165 = 0.835$.

$\boxed{P(X \geq 1) = 0.835}$

4. If Poisson variable 'X' is such that, $P(X=1) = 2P(X=2)$ find $P(X=0)$ and var of X

Soln:

The p.mf of Poisson Distribution is,

$$P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2 \ldots$$

Given:

$P(X=1) = 2P(X=2)$

$$P(X=1) = \frac{e^{-\lambda}\lambda}{1} = \cancel{e^{-\lambda}}\lambda^{\cancel{1}} = \frac{\cancel{2}\,\cancel{e^{-\lambda}}\,\lambda^2}{\cancel{2}}$$

$\lambda^{\cancel{1}} = \lambda^2$

$\lambda^2 = \lambda$ $\quad \boxed{\lambda = 1}$

The p.mf $P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-1} \cdot (1)^x}{x!}$ $x = 0, 1, 2 \ldots$

i) $P(X=0) \Rightarrow \frac{e^{-1}(1)^0}{0} = \boxed{0.367 = P(X=0)}$

ii) Var of $X = \lambda \Rightarrow \boxed{\lambda = 1}$ = Var of X.

5. The average No. of Traffic accidents on a certain section, of a highway is 2 per week Assume that the No. of accidents follows a Poisson distribution. Find the Probability that

i) No Accident in a week

ii) Atmost 3 accident in 2 week Period

Soln: Let X be the No of Accidents,

$\boxed{\lambda = 2}$ per week.

In P.D, The p.mf $P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-2}(2)^x}{x!}$ $x = 0, 1, 2, 3$.

$P(X=0) = \frac{e^{-2}(2)^0}{0}$ $\boxed{0.135 = P(X=0)}$ $e^{-1}(2)$ e^{-}

During a two week period, the average No. of accidents would be 4.

$\frac{e^{-2}(2)^x}{x!}$

ii) P(atmost 3 Accidents in a 2 Week Period) =

$= P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3).$

$= 0.135 + 0.27 + 0.27 + 0.181.$ $\frac{e^{-2}(2)^2}{2!}$ $\underline{0.54}$

$\boxed{P(X \leq 3) = 0.856}$

$\boxed{N \to \text{large} \downarrow \text{Poisson}}$

<u>Geometric Distribution</u>: $\boxed{Mgf = \frac{Pe^t}{1-qe^t}}$ $\boxed{Mean = \frac{1}{P}}$ $\boxed{\text{Var of } X = \frac{q}{P^2}}$

A Random Variable X is said to have a Geometric Distribution with parameter 'p' if the p.mf is given by,

$\boxed{P(X=x) = q^{x-1}p, \quad x = 1, 2, \ldots}$ $\boxed{P(X=x)\ q^x . P \quad Mgf = \frac{P}{1-qe^t}.}$

where $q = 1-p$, $p+q=1$.

(1). Find Mean, variance, Mgf of Geometric Distribution.

<u>Soln</u>:

The p.mf of G.D is,

$P(X=x) = q^{x-1}P \quad ; x = 1, 2, 3 \ldots$

<u>Mgf</u>: $= M_X(t) = E[e^{tx}] = \sum_{x=1}^{\infty} e^{tx}p(x_i)$

$$= \sum_{x=1}^{\infty} e^{tx} . q^{x-1} . P$$

$$= \frac{P}{q} \sum_{x=1}^{\infty} e^{tx} . q^x .$$

$$= \frac{P}{q} \sum_{x=1}^{\infty} (e^t . q)^x$$

$$= \frac{P}{q} \left[(e^t q)^1 + (e^t q)^2 + (e^t q)^3 + \ldots \right.$$

$$= \frac{P}{q} e^t q \left[1 + (e^t q)^1 + (e^t q)^2 + \ldots \right.$$

$$\frac{p}{q} \cdot qe^t \left[\frac{1}{1-qe^t}\right] = pe^t \left[\frac{1}{1-qe^t}\right] \Rightarrow \boxed{\frac{pe^t}{1-qe^t} = M_X(t)}$$

$M_X(t) \; \frac{u}{v} = \frac{vu' - uv'}{v^2}$

$$M_x'(t) = \frac{(1-qe^t)(pe^t) - (pe^t)(-qe^t)}{(1-qe^t)^2}$$

$$= pe^t - \cancel{pq}e^{2t} + pq\,\cancel{e^{2t}} = \frac{pe^t}{(1-qe^t)^2} = M_x'(t)$$

$$M_x'(t) = \frac{pe^t}{(1-qe^t)^2}$$

$$M_x''(t) = \frac{(1-qe^t)^2(pe^t) - (pe^t)\,2(1-qe^t)(qe^t)}{(1-qe^t)^4}$$

$$= \frac{(1-qe^t)^2(pe^t) + 2pe^{2t}q^2e^{3t}\;2pqe^{2t} - 2pq^2e^{3t}}{(1-qe^t)^4}$$

$$E(X) = M_x'(0) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \boxed{\frac{1}{p} = E(X)}$$

$$E[X^2] = M_x''(0) = \frac{(1-q)^2 p - 2p(1-q)(-q)}{(1-q)^4}$$

$$= \frac{p^3 + 2p^2q}{p^4} = \frac{p^2[p+2q]}{p^4}$$

$$= \frac{p+2q}{p^2} = \frac{p+q+q}{p^2} = \boxed{\frac{1+q}{p^2} = E(X^2)}$$

$$\text{Var of }(X) = E[X^2] - (E[X])^2$$

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

Another form:

If X denotes the Number of Failures before the 1st Success then,

$$\boxed{P(X=x) = q^x p \Rightarrow x = 0, 1, 2 \ldots}$$

$$\boxed{M_X(t) = \frac{p}{1-qe^t}}, \quad \boxed{E(X) = \frac{q}{p}}, \quad \boxed{\text{Var of }(X) = \frac{q}{p^2}}$$

Con → exponential
Dis → Geom

statement: 2m

Memory less Property:

of Geometric Distribution:

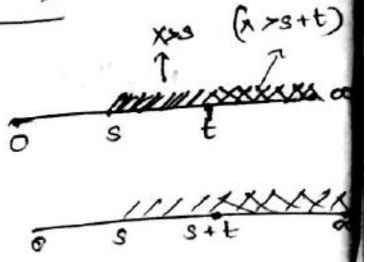
If X is a Random Variable, with the Geometric Distribution, then,

M.L.P. (G, E)

$$P(X > s+t / X > s) = P(X > t)$$

$$P(X>s+t / X>s) = \frac{P(X > s+t \cap X > s)}{P(X > s)}$$

$P(A/B)$



$$= \frac{P(X > s+t)}{P(X>s)} \quad \text{--- (1)}$$

Let $P(X > k) = \sum_{x=k+1}^{\infty} q^{x-1} \cdot p$

continuous: $P(x>2) = P(x=3) + P(x=4) + \ldots$ $\sum_{x=3}^{\infty}$

$= (q^k p) \, p \, q$

$= P[e^t + e^{t-}$

$\Rightarrow P(q^k + q^{k+1} + q^{k+2} + \ldots$

$= Pq^k[1 + q + q^2 + \ldots] = Pq^k\left[\frac{1}{1-q}\right]$

$= \frac{\cancel{p} q^k}{\cancel{p}} = q^k$

$$P(X>k) = q^k \quad \text{--- (2)}$$

using (2) in (1),

$$P(X>s+t / X>s) = \frac{q^{s+t}}{q^s} = \frac{\cancel{q^s} \cdot q^t}{\cancel{q^s}} = q^t.$$

$$P(X>s+t / X>s) = P(X>t) \text{ (using (1))}.$$

A training solider shoots a target in a independent fashion. if the probability that shoot on any one shoot is 0.8 is,

i) What is the Probability that the target would be? [1st hits at the 6th attempt]

ii) What is the Probability that it takes less than 5 shoots.

Soln:

In geometric distribution,

$\boxed{p = 0.8}$, $q = 1 - 0.8$ $\boxed{q = 0.2}$

The p.m.f $P(X=x) = q^{x-1} p$, $x = 1, 2, 3 \ldots$

$= (0.2)^{x-1}(0.8)$

i) p(target is hit on 6^{th} attempt) = $P(X=6)$

$= (0.2)^5 (0.8) = 2.56 \times 10^{-4}$

ii) p(less than 5 shoots) = $P(X<5)$

$= P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= (0.2)^0(0.8) + (0.2)^1(0.8) + (0.2)^2(0.8) + (0.2)^3(0.8)$

$= 0.8[1 + 0.2 + 0.04 + 0.008]$

$= 0.998$

8m

(Q) A dice is tossed until 6 appear, what is the probability that, it must be tossed more than 5 times?

Soln:

Let x be the no. of tosses required to get, the first 6.

P(getting 6) = $\boxed{p = 1/6}$, $q = 1 - 1/6 = \boxed{q = \frac{5}{6}}$

The p.m.f $P(X=x) = q^{x-1} p$, $x = 1, 2, 3 \ldots$

$= \left(\frac{5}{6}\right)^{x-1}\left(\frac{1}{6}\right)$ $x = 1, 2, 3 \ldots$

P(more than 5 tries) $= P(X>5) = 1 - P(X \le 5)$

$= 1 - [P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)]$

$= 1 - \left[\left(\frac{5}{6}\right)^0\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right)\right]$

$= 1 - \left(\frac{1}{6}\right)\left[1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^4\right]$

$= 1 - \left(\frac{1}{6}\left[\frac{1}{1 - 5/6}\right]\right) = \frac{5}{6}\left[\frac{6}{1}\right]$

$= \cancel{1} - \cancel{\frac{1}{6}}\left[\frac{6}{1}\right] \Rightarrow \boxed{0.401 = P(X>5)}$

Another method:

Property:

$P(X>k) = q^k$

$P(X>5) = \left(\frac{5}{6}\right)^5$

$= 0.401$

(8) If the probability that if the target is destroyed on any one shot is 0.5, what is the probability that it could be destroyed on 6^{th} attempt.

Soln: $\boxed{p = 0.5}$ $q = 1 - 0.5$ $\boxed{q = 0.5}$

The p.mf G.D is $P(X=x) = q^{x-1} p \quad x = 1, 2, 3.$

$P(X=x) = (0.5)^{x-1}(0.5)$

P(

P(6^{th} attempt) $P(X=6) = (0.5)^5 (0.5)$

$= 0.03125$

$\boxed{P(X=6) = 0.0156}$

(8) If $M_X(t) = (5-4e^t)^{-1}$ Find the probability (mgf) of P (5 (or) 6)

Soln: $M_X(t) = \frac{1}{(5-4e^t)} = \frac{1}{5\left(1 - \frac{4}{5}e^t\right)} = \frac{1/5}{\left(1 - \frac{4}{5}e^t\right)}$

The mgf of G.D is $M_X(t) = \frac{p}{1 - qe^t}$.

$\boxed{p = 1/5}$ $\boxed{q = \frac{4}{5}}$

$\therefore$ p.mf $P(X=x) = q^x p. \quad x = 0, 1, 2, 3 \ldots$

$= P(X=5 \text{ (or) } 6) = P(X=5) + P(X=6)$

$= \left(\frac{4}{5}\right)^5 \left(\frac{1}{5}\right) + \left(\frac{4}{5}\right)^2 \left(\frac{1}{5}\right)$

$\boxed{P(X=5 \text{ (or) } 6) = 0.117.}$

Uniform Distribution (density function) continuous

A continuous Random Variable, x, defined in the interval (a, b) is said to follow an uniform distribution,

$$f(x) = \frac{1}{b-a}, \quad a < x < b.$$

Find Mean, Variance and mgf of Uniform distribution,

soln:

The p.d.f of U.D is $f(x) = \frac{1}{b-a}$; $x < a < b$

$$\text{Mean} = E[x] = \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b \left(\frac{1}{b-a}\right) x\,dx$$

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

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

$$\boxed{\text{Mean} = \frac{b+a}{2}}$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx \Rightarrow \int_a^b x^2 \cdot \frac{1}{b-a}\,dx \Rightarrow \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$

$$\Rightarrow \frac{1}{3(b-a)}[b^3 - a^3] \Rightarrow \frac{(b^3-a^3)}{3(b-a)} \Rightarrow \frac{\cancel{(b-a)}(a^2+b^2+ab)}{3\cancel{(b-a)}}$$

$$\Rightarrow \frac{a^2+b^2+ab}{3}$$

$$\text{Var of } x = E[x^2] - (E[x])^2.$$

$$= \frac{a^2+b^2+ab}{3} - \frac{(b+a)^2}{4} = \frac{a^2+b^2+ab}{3} - \frac{b^2+a^2+2ab}{4}$$

$$= \frac{\cancel{a^2+b^2+ab-4b^2-4a^2-2ab}}{12} \Rightarrow \frac{4a^2+4b^2+4ab-3b^2-3a^2-6ab}{12}$$

$$= \frac{\cancel{ab}}{\cancel{12}} = \frac{b^2+a^2-2ab}{12}$$

$$\boxed{\text{Var of } x = \frac{(b-a)^2}{12}}$$

$$M_X(t) = mgf = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\,dx.$$

$$= \int_a^b e^{tx} \cdot \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b \Rightarrow M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}.$$

$$\boxed{Mgf = \frac{e^{bt} - e^{at}}{t(b-a)}}$$

less < 5.

(b-a) (0-30) / (10-15) (25-30).

>10 (0-5) (15-20)

Q. Buses arrived at specified stop at 15 min interval, starting at 7 A.M. (ie) they arrive at 7.15, 7.30, 7.45 . . . AM. if a passenger arrive at a stop at a Random time is uniformly Distributed between 7 and 7.30 AM. Find the probability that, he waits for

i) less than 5 mins for a bus.

ii) more than 10 mins for a bus.

Soln:

Let X denote the time taken a passenger arrive bt, 7 - 7.30 AM

In U.D the pdf,

$$f(x) = \frac{1}{b-a} = \frac{1}{30-0} = \frac{1}{30}$$

i) passenger waits less than 5 mins

(ie) he arrived bt, 7.10 - 7.15 AM, (or) 7.25 - 7.30 AM.

P (to wait less than 5 mins) = P (10 < x < 15) + P (25 < x < 30)

$$= \int_{10}^{15} \frac{1}{30}\,dx + \int_{25}^{30} \frac{1}{30}\,dx.$$

right short way

60 (5-15) + (35-45)

after 15 mins

ii) (15-30) (45-60)

$= \frac{1}{30}[x]_{10}^{15} + \frac{1}{30}[20 - 25] \Rightarrow \frac{1}{30}[5] + \frac{1}{30}[5]$

$= \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \boxed{\frac{1}{3}}$

ii) passenger waits for more than 10 mins, he arrived bt (7 - 7.05 AM (or) 7.15 - 7.20 AM

(10 mins) $\downarrow$ (0 - 5) (15 - 18)

P(he waits for more than 10 mins)

$= P(0 < x < 5) + P(15 < x < 20)$

$= \int_0^5 \frac{1}{30}\,dx + \int_{15}^{20} \frac{1}{30}\,dx$

$= \frac{1}{30}[5] + \frac{1}{30}[5] = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \boxed{\frac{1}{3}}$

(a). If x is uniformly distributed, (0, 10) calculate the probability that

i) $P(x < 3)$ ii) $P(x > 6)$ iii) $P(3 < x < 8)$

soln:

P.df of U.D $= \frac{1}{b-a}$; $a < x < b$;

$= \frac{1}{10-0} = \frac{1}{10}$; $0 < x < 10$.

i) $P(x < 3)$: $= \int_0^3 \frac{1}{10}\,dx = \frac{1}{10}[x]_0^3 = \frac{3}{10}$

ii) $P(x > 6)$: $= \int_6^{10} \frac{1}{10}\,dx = \frac{1}{10}\left[x\right]_6^{10} = \frac{1}{10}[4] = \frac{4}{10} = \frac{2}{5}$

iii) $P(3 < x < 8) = \int_3^8 \frac{1}{10}\,dx \Rightarrow \frac{1}{10}[x]_3^8 = \frac{1}{10}[5] = \frac{1}{2}$ //.

(b). A Random Variable X is Uniformly distributed over (-3, 3) compute,

i) $P(x < 2)$ ii) $P(|x| < 2)$ iii) $P(|x-2| < 2)$.

find K for which $P(x > K) = \frac{1}{3}$.

Soln: The pdf of $f(x) = \frac{1}{b-a}$, $a < x < b$;

$$f(x) = \frac{1}{3-(-3)} = \frac{1}{6}, \quad -3 < x < 3.$$

i) $P(x<2) = \int_{-3}^{2} \frac{1}{6}\, dx \Rightarrow \frac{1}{6}[2+3] = 5/6.$

ii) $P(|x|<2). = \int_{-2}^{2} \frac{1}{6}\, dx \Rightarrow \frac{1}{6}[4] = \frac{2}{3}.$

iii) $P(|(x-2)|<2) = P(-2 < x-2 < 2) = P(0 < x < 4)$ $(-3, 3)$

$$\Rightarrow \int_{0}^{3} f(x)\, dx \Rightarrow \left[\frac{1}{6} x\right]_0^3 \Rightarrow \frac{1}{6}[3] = \frac{1}{2}$$

iv) $P(x>K) = \frac{1}{3}.$

$$\Rightarrow \int_{K}^{3} \frac{1}{6}\, dx = \frac{1}{3} \Rightarrow \left.\frac{1}{6} x\right]_K^3 = \frac{1}{3}$$

$$\Rightarrow 3-K = \frac{1}{3}(6)$$

$$3-K = 2 \qquad \boxed{K=1}$$

(9). The no. of P.C of sold, daily at the computer world, it is uniformly distributed with minimum of 2000 Pc and maximum of 5000 PC. Find the,

i) the Probability that daily sales, will fall between 2500 and 3000.

ii) what is the probability that the computer will sell, at least 4000 PC's.

iii) what is the probability that the computer will exactly sell 2500 PC's.

Soln: The p.d.f of U.D $f(x) = \frac{1}{b-a}$; $a < x < b$

$= \frac{1}{5000-2000}$; $2000 < x < 5000$

$f(x) = \frac{1}{3000}$.

i) $P(2500 < X < 3000) = \int_{2500}^{3000} \frac{1}{3000} dx$

$= \frac{1}{3000}\left[x\right]_{2500}^{3000} = \frac{1}{3000}[500]$

$= \boxed{\frac{1}{6}}$.

ii) $P(X \geq 4000) = 1 - P(X < 4000)$.

$= 1 - \int_{2000}^{4000} \frac{1}{3000} dx \Rightarrow$

$= \int_{4000}^{5000} \frac{1}{3000} dx \Rightarrow \frac{1}{3000}[x]_{4000}^{5000} = \frac{1}{3000}[1000]$

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

iii) P(exactly sell 2500 pcs) $= P(X = 2500) = \boxed{0}$

$P(X = 2500)$

continuous

$\int_{2500}^{2500} \Rightarrow 0$

(8). If X is uniformly distributed with mean 1, and variance 4/3. Find $P(X \leq 0)$.

Soln: In U.D, Mean $= \frac{b+a}{2} = 1$ $\quad b+a = 2$. —(1)

Variance $X = \frac{(b-a)^2}{12} = \frac{4}{3} \Rightarrow (b-a)^2 = 16$.

$b - a = 4$. —(2).

(1) + (2),

$b + a = 2$

$b - a = 4$

$2b = 6$

$\boxed{b = 3}$

$\Rightarrow 3 - a = 4$

$-a = 4 - 3$

$\boxed{a = -1}$

continuous >, ≥ Same.

$\therefore$ The p.d.f of U.D $f(x) = \frac{1}{b-a}$; $a < x < b$

$f(x) = \frac{1}{3+1} = \frac{1}{4}$ $\quad -1 < x < 3$,

$P(X \leq 0) = \int_{-1}^{0} \frac{1}{4} dx \Rightarrow \frac{1}{4}[x]_{-1}^{0} \Rightarrow \boxed{\frac{1}{4} = P(X \leq 0)}$

$M_x(t) = (1-t)^{-\lambda}$

$M'_x(t) = -\lambda (1-t)^{-\lambda-1}(-1)$

$= \lambda (1-t)^{-\lambda-1}$

$M''_x(t) = (\lambda)(-\lambda-1)(1-t)^{-\lambda-2}(-1)$

$= -(\lambda)(\lambda+1)(1-t)^{-\lambda-2}(-1)$

$= \lambda(\lambda+1)(1-t)^{-\lambda-2}$

$M'_x(0) = \lambda, \qquad M''_x(0) = \lambda^2+\lambda.$ $\qquad \lambda e^{-\lambda x} \quad x>0$

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

$= \cancel{\lambda^2}+\lambda - \cancel{\lambda^2}$

$\boxed{\text{Var of } x = \lambda.}$ $\qquad \boxed{\text{Mean} = \lambda}$

$\boxed{Mgf = \frac{\lambda}{\lambda - t}}$ $\boxed{\text{Mean} = \frac{1}{\lambda}}$ $\boxed{\text{var} = \frac{1}{\lambda^2}}$

Exponential Distribution: (λ)

A continuous random variable X, defined in the interval $(0,\infty)$ is said to follow exponential distribution, with parameter λ with p.d.f,

$$\boxed{f(x) = \lambda e^{-\lambda x}, \quad x>0,\ \lambda>0}$$

Q. Find Mean, Variance, Mgf of Exponential distribution.

Soln:

The Mgf $M_x(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx.$

$= \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x}\, dx.$

$= \lambda \int_0^{\infty} e^{-(\lambda - t)x}\, dx$

$= \lambda \left[\frac{e^{-(\lambda-t)x}}{-(\lambda-t)} \right]_0^{\infty}$

$= \lambda \left[-\frac{1}{\lambda-t}(0) + \frac{1}{\lambda-t}(1) \right]$

$\boxed{Mgf = \frac{\lambda}{\lambda - t}.}$ $\Rightarrow$ $\boxed{\lambda(\lambda-t)^{-1} = Mgf}$

$M'_x(t) = -2\lambda(\lambda-t)^{-3}(-1)$

$M'_x(t) = \lambda(-1)(\lambda-t)^{-2}(-1)$

$= \lambda(\lambda-t)^{-2}$.

$M''_x(t) = -2\lambda(\lambda-t)^{-3}(-1)$

$= 2\lambda(\lambda-t)^{-3}$.

$M'_x(0) = \lambda.\lambda^{-2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} = E[x]$

$M''(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2} = E[x^2]$

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

$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

$\boxed{\text{Var of } x = \frac{1}{\lambda^2}}$ $\boxed{\text{Mean} = \frac{1}{\lambda}}$

Memory less property of Exponential (λ) Distribution: E.F:

continuous

If x is exponentially distributed, with parameter λ with any +ve integer, then

$$\boxed{P(x > s+t / x > s) = P(x > t)}$$

The p.d.f of E.D is,

$f(x) = \lambda.e^{-\lambda x}, \quad x > 0$

Let $P(X > k) = \int_k^\infty \lambda . e^{-\lambda x}\, dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_k^\infty$

$= [e^{-\lambda k}]$ —— (1).

$P(x > s+t / x > s) = \frac{P(x > s+t \cap x > s)}{P(x > s)} = \frac{P(x > s+t)}{P(x > s)}$.

$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = \frac{\cancel{e^{-\lambda s}} . e^{-\lambda t}}{\cancel{e^{-\lambda s}}}$

$= e^{-\lambda t} = P(x > t)$ —— using (1), The time (hrs) ~~required to~~ repair a machine X ~~is exceeds~~ is exponential distributed with parameter $\lambda = 1/2$. i) what is the Probability that the required repaired time exceeds 2 hrs. $P(X>2)$

ii) What is conditional Probability (M.L.P) the a repair ~~that required~~ time takes the ~~exceeds~~ at least 10 hrs is given that Given that, iii) duration exceeds 9 hrs. $P(X>10/X>9)$ $s+t$ s $= P(X>1)$ t

Soln:

The p.d.f of Exponential distribution is,

$$f(x) = \lambda e^{-\lambda x}, \ x>0.$$

Given $\lambda = 1/2 \Rightarrow \frac{1}{2} e^{-x/2}, \ x>0.$

i) $P($repair time x is exceeds 2 hrs$) = P(X>2)$.

$$\Rightarrow \int_2^\infty \frac{1}{2} e^{-x/2} dx = \frac{1}{2}\int_2^\infty e^{-x/2} dx.$$

$$\Rightarrow \frac{1}{2} \left[\frac{e^{-x/2}}{-1/2}\right]_2^\infty \Rightarrow \left[0 - (-e^{-1})\right] = e^{-1} \Rightarrow \frac{1}{e} = 0.3676.$$

ii) $P(X>10/X>9) = P(X>1)$. By:

$\underset{9+1}{10}$, $\underset{s}{9}$

By memoryless Pro:

$P(X>s+t/X>s) = P(X>t)$

$\underset{10}{s+t}$, $\underset{9}{s}$, $\underset{10-9=1}{t}$

$$= \int_1^\infty \frac{1}{2} e^{-x/2} dx.$$

$$= \frac{1}{2}\left[\frac{e^{-x/2}}{-1/2}\right]_1^\infty \Rightarrow 0 - (-e^{-1/2})$$

$$= 0.6$$

(Q in G.D)

The daily consumption of milk in a city in excess of 20,000 litres is approximately exponentially distributed with mean 3000 litres. The city has daily stock of 35,000 litres. What is the probability that of two days, selected at random, the stock is insufficient for the days.

Soln:

Let x denotes excess of consumption of milk.

y denotes consumption of the milk

In exponential distribution,

Given; Mean $= \frac{1}{\lambda} = \frac{1}{\cancel{3000}} = 3000 = \frac{1}{\lambda}$

$$\boxed{\lambda = \frac{1}{3000}}$$

In exp. Dis,

$$Pdf = f(x) = \lambda e^{-\lambda x} = \frac{1}{3000} e^{-x/3000}$$

P(insufficient stock for 1 day) $= P(y > 35000)$

$= P((x + 20,000) > 35000)$

$= P(x > 35000 - 20,000)$

$= P(x > 15000)$.

$$= \int_{15000}^{\infty} \frac{1}{3000} e^{-x/3000} \, dx$$

$$= \frac{1}{\cancel{3000}} \left[\frac{e^{-x/3000}}{-1/\cancel{3000}} \right]_{15000}^{\infty}$$

$$= \left[0 - (-e^{-5})\right] \Rightarrow e^{-5} \Rightarrow 6.73 \times 10^{-3}$$

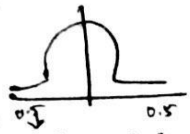
P(insufficient stock for both days] $= e^{-5} \times e^{-5}$

$= e^{-10}$

$= \boxed{4.53 \times 10^{-5}}$

Normal Distribution.

$s > 30 \to L.S$, $s < 30 \to S.S$.



0.5 0.5

1 Tail test 2 tail test.

The p.d.f of Normal Distribution, is,

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

$-\infty < x < \infty$.

$-\infty < \sigma < \infty$.

$\mu > 0$.

$\boxed{\begin{array}{l} \text{Mean} = \mu \\ \text{Var} = \sigma^2 \end{array}}$

$\mu \to$ mean. Parameter (μ, σ)

$\sigma \to$ Variance.

S.D.

Q. Find mean, Variance, mgf of Normal Distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Soln:

The pdf of N.D is,

$$\text{mgf}: M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} e^{tx}\,dx.$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Let $\boxed{z = \frac{x-\mu}{\sigma}}$

$\sigma z = x - \mu$ $\qquad x = -\infty, z = -\infty$

$\sigma dz = dx.$ $\qquad x = \infty, z = \infty.$

$x = \sigma z + \mu.$

$$= \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}} \int_{-\infty}^{\infty} e^{t(\sigma z + \mu)} \cdot e^{-\frac{1}{2}z^2} \cancel{\sigma}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\sigma z t} \cdot e^{-z^2/2}\, dz$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(\frac{-z^2 + 2\sigma z t}{2}\right)} dz.$$

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(\frac{-z^2 + 2\sigma z t - \sigma^2 t^2}{2}\right)}.$$

$$= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{z^2 - 2\sigma t z + \sigma^2 t^2}{2}\right)} dz.$$

$$\Rightarrow \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2} dz.$$

$$\Rightarrow \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{z-\sigma t}{\sqrt{2}}\right)^2} dz.$$

<u>sub:</u> $u = \frac{z - \sigma t}{\sqrt{2}}$ $\Rightarrow \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\sqrt{\cancel{2}\pi}} \int_{-\infty}^{\infty} e^{-u^2} \cancel{\sqrt{2}}\, du.$

$\frac{du}{dz} = \frac{1}{\sqrt{2}}$

$dz = \sqrt{2}\,du.$ $= \frac{e^{\mu t + \frac{\sigma^2 t^2}{2}}}{\cancel{\sqrt{\pi}}} \cancel{\sqrt{\pi}}$

$$\boxed{Mgf = e^{\mu t + \frac{\sigma^2 t^2}{2}}}$$

$$M_x'(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \left[\mu + \frac{\cancel{2}\sigma^2 t}{\cancel{2}}\right]$$

$$M_x''(t) = \sigma^2 \left[e^{\mu t + \frac{\sigma^2 t^2}{2}}\right] + (\mu + \sigma^2 t)\, e^{\mu t + \frac{\sigma^2 t^2}{2}} [\mu + \sigma^2 t]$$

$E[x] = M_x'(0) = \mu.$

$E[x^2] = M_x''(0) = \mu^2 + \sigma^2.$

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

$= \cancel{\mu^2} + \sigma^2 - \cancel{\mu^2}$

$$\boxed{Var\ of\ x = \sigma^2}$$

If x is a normal variate with mean $\mu = 30$, $\sigma = 5$, find,

i) $P(20 \le x \le 40)$ ii) $P(x \ge 45)$ iii) $P(|x-30| > 5)$.

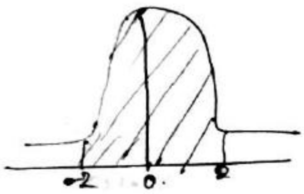
Given: $\mu = 30$, $\sigma = 5$.

Normal distribution,

Normal Variate $= z = \dfrac{x-\mu}{\sigma}$

$$z = \frac{x-30}{5}.$$

i) $P(20 \le x \le 40) = P\left(\dfrac{20-30}{5} \le z \le \dfrac{40-30}{5}\right)$



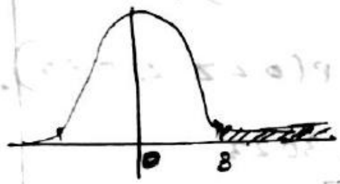
$= P(-2 \le z \le 2)$

$= 2P(0 \le z \le 2)$

$= 2(0.4772)$

$$\boxed{P(20 \le x \le 40) = 0.9544.}$$

ii) $P(x \ge 45) = P\left(z \ge \dfrac{45-30}{5}\right) = P(z \ge 3)$.



$= \cancel{1-P} = 0.5 - P(0 \le z \le 3)$.

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

$= 0.5 - 0.4895$.

$$\boxed{P(x \ge 45) = 0.0105.}$$

iii) $P(|x-30| > 5) = 1-[P(-5 < x-30 < 5)]$

$= 1-[P(25 < x < 35)]$

$= 1-\left[P\left(\dfrac{25-30}{5} < z < \dfrac{35-30}{5}\right)\right]$

$= 1-[P(-1 < z < 1)]$

$= 1-[2P(0 < z \le 1)]$

$= 1-[2 \times 0.3413]$

$\boxed{P(|x-30| > 5) = 1-[0.6826].} = 0.3174$

(8). An electric firm manufactures bulb that has life before burn out (i.e.) Normally distributed with mean 800 hrs. At standard deviation 40 hrs. find,

i) Probability a bulb burns more than 8,34 hrs.

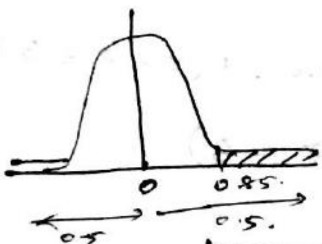
ii) Probability bulb burns between 7,78 and 834 hrs.

Soln: $\mu = 800$, $\sigma = 40$

Normal variant $z = \frac{x-\mu}{\sigma} = \frac{x-800}{40}$

P (a bulb burns more than 8,34 hrs)

$= P(x \geq 834) = P\left(z \geq \frac{834-800}{40}\right)$



$= P(z \geq 0.85)$.

$= P(0.5 - P(0 < z < 0.85)$.

$= 0.5 - 0.3023$

$\boxed{P(x \geq 834) = 0.1977.}$

ii) P(bulbs burns bt 778 & 834 hrs)

$\Rightarrow P(778 \leq x \leq 834) = P(-0.55 \leq z \leq 0.85)$.

$\Rightarrow P\left(\left(\frac{778-800}{40}\right) \leq z \leq \frac{834-800}{40}\right) = P(0 \leq z \leq 0.55) + P(0 \leq z \leq 0.85)$

$= 0.2088 + 0.3023$

$= \boxed{0.5111.}$

UNIT-2

TWO DIMENSIONAL RANDOM VARIABLES

DEFINITION:

Let S be a sample space and let $X = X(S)$ and $Y = Y(S)$ be two functions, Each assigning a real Number. Each outcome set S, then (X, Y) is a two dimensional Random variable.

Two-Dimiensional Discrete Random variable:

If the possible values of X, Y are finite, then (X, Y) is called a two dimensional discrete Random variable. And it can be represented by $P(x_i, y_j)$.

Joint Probability Mass Function: (p.m.f)

i) $P(x_i, y_j) \geq 0$

ii) $\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) = 1$

Marginal Probability Distribution:

X \ Y	Y_1	Y_2	...	Y_m	$P(X=X_i)$
X_1	P_{11}	P_{12}		P_{1m}	$P(X=X_1)$
X_2	P_{21}	P_{22}		P_{2m}	$P(X=X_2)$
⋮					⋮
X_n	P_{n1}	P_{n2}		P_{nm}	$P(X=X_n)$
$P(Y=y_j)$	$P(Y=y_1)$	$P(Y=y_2)$		$P(Y=y_n)$	1

The Marginal probability function if X is,

$P(X=X_1), P(X=X_2) \ldots P(X=X_n)$.

The Marginal probability function if Y is,

$P(Y=y_1), P(Y=y_2) \ldots P(Y=y_m)$.

Conditional Probability function of X:

conditional probability function of x given $y = y_j$

$$P\left(\frac{x=x_i}{y=y_j}\right) = \frac{P(x=x_i \cap y=y_j)}{P(y=y_j)} = P_{ij}/P_{\cdot j}$$

Conditional probability function of Y:

conditional probability function of y given $x = x_i$

$$P\left(\frac{y=y_j}{x=x_i}\right) = \frac{P(y=y_j \cap x=x_i)}{P(x=x_i)} = P_{ji}/P_{i\cdot}$$

Conditional density function x on y and y on x:

x on y; $f(x/y) = \frac{f(x,y)}{f(y)}$

y on x; $f(y/x) = \frac{f(x,y)}{f(x)}$

ii) Two Dimensional Continous Random Variable:

If (x,y) can assume all the values in a specified region R in (x,y) plane, then (x,y) is called as Two Dimensional continuous Random variable.

Join Probability density Function: (p.d.f)

i) $f(x,y) \geq 0$

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

Marginal p.d.f of x is given by y:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,d\cancel{x}y$$

Marginal p.d.f of y is given by x:

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

Independent:

If x and y are independent,

i) $f(x) \cdot f(y) = f(x,y)$ [continuous case]

$P(X=i).P(X=j) = P_{ij}$ [Discrete case]

Note:

$$P(a_1 < x < b_1, a_2 < y < b_2) = \int_{a_2}^{b_2}\int_{a_1}^{b_1} f(x,y)\,dx\,dy$$

Joint probability Distribution Function:

$$F(x) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy \text{ [continuous case]}$$

$$F(x) = P(x \le x), y \le x) \text{ [Discrete case]}$$

$$= \sum_{x_i < x}\sum_{y_i < y} P(x_i, y_i)$$

Discrete case:

1. From the following distribution of (x,y),

$x \backslash y$	1	2	3	4	5	6
0	0	0	$1/32$	$2/32$	$2/32$	$3/32$
1	$1/16$	$1/16$	$1/8$	$1/8$	$1/8$	$1/8$
2	$1/32$	$1/32$	$1/64$	$1/64$	0	$2/64$

i) Find $P(x \le 1), P(y \le 3)$

ii) Find $P(x \le 1, y \le 3)$

iii) Find $P(x \le 1 / y \le 3)$

iv) Find $P(y \le 3 / x \le 1)$

v) Find Marginal Distribution of x.

vi) Find Marginal Distribution of y.

vii) Find conditional distribution of x, given $y = 2$.

viii) Examine x and y are independent.

ix) Find $E(y-2x)$

Solution:

Given:

X \ Y	1	2	3	4	5	6	$P(X=x_i)$
0	0	0	1/32	2/32	2/32	3/32	$P(X=0)=8/32$
1	1/16	1/16	1/8	1/8	1/8	1/8	$P(X=1)=10/16$
2	1/32	1/32	1/64	1/64	0	2/64	$P(X=2)=8/64$
$P(Y=y_j)$	$P(Y=1)=3/32$	$P(Y=2)=3/32$	$P(Y=3)=11/64$	$P(Y=4)=13/64$	$P(Y=5)=6/32$	$P(Y=6)=16/64$	1

i) Marginal Distribution of X:

$P(X=0)=8/32$; $P(X=1)=18/16$; $P(X=2)=8/64$

ii) Marginal Distribution of Y:

$P(Y=1)=3/32$; $P(Y=2)=3/32$; $P(Y=3)=11/64$; $P(Y=4)=13/64$; $P(Y=5)=6/32$;

$P(Y=6=16/64)$.

iii) $P(X\le 1) = P(X=0)+P(X=1)$

$= 8/32 + 10/16$

$= 28/32$

$= 7/8$

iv) $P(Y\le 3) = P(Y=1)+P(Y=2)+P(Y=3)$

$= 3/32 + 3/32 + 11/64$

$= 23/64$

v) $P(X\le 1 / Y\le 3) = \dfrac{P[X\le 1, Y\le 3]}{P(Y\le 3)}$

$= \dfrac{9/32}{23/64}$

$= 18/23$

vi) $P(X\le 1, Y\le 3) = 0+0+1/32+1/16+1/16+1/8$

$= 1/32 + 2/16 + 1/8$

$= 9/32$

vii) $P(y \le 3 / x \le 1) = \dfrac{P(y \le 3).P(x \le 1)}{P(x \le 1)}$

$= \dfrac{9/\cancel{32}}{28/\cancel{32}}$

$= 9/28$

viii) $P[x+y \le 4] \Rightarrow P_{01} + P_{02} + P_{03} + P_{04} + P_{11} + P_{12} + P_{13} + P_{21} + P_{22}$

$= 0+0+1/32 + 2/32 + 1/16 + 1/16 + 1/8 + 1/32 + 1/32$

$= 13/32$

ix) conditional Distribution of x, given y=2.

$P(x=0/y=2) = \dfrac{P_{02}}{P(y=2)} = 0$

$P(x=1/y=2) = \dfrac{P_{12}}{P(y=2)} = \dfrac{1/16}{3/32} = 2/3$

$P(x=2/y=2) = \dfrac{P_{22}}{P(y=2)} = \dfrac{1/32}{3/32} = 1/3$

x) $p(x=1) \times p(x=2) = P_{12}$ (Property).

$10/16 \times 8/64 \neq 1/16$

$\therefore$ x and y are not independent

xi) $E(y-2x) = E(y) - 2E(x)$

$E(x) = \sum x_i p(x_i)$

$= 0 \times 8/32 + 1 \times 10/16 + 2 \times 8/64$

$= 0 + 10/16 + 16/24$

$= 5/8 + 4/6$

$= 56/64$

$= 7/8$

$E(y) = \sum y_j P(y_j)$

$= 1\times 3/32 + 2\times 3/32 + 3\times 11/64 + 4\times 13/64 + 5\times 6/32 + 6\times 16/64$

$= 3/32 + 6/32 + 33/64 + 52/64 + 30/32 + 96/64$

$= 239/64$

$E(y-2x) = 239/64 - 2(7/8)$

$= 147/64$

2. Joint probability function of x, y is given by, $p(x,y) = K(2x+3y)$ $x = 0,1,2$, $y = 1,2,3$

i) Find marginal probability distribution of x and y.

ii) Find probability distribution of x+y.

iii) $P((x+y) > 3)$

iv) Find all conditional distributions.

Sol:

X \ Y	1	2	3	$P(X=x_i)$
0	3K	6K	9K	18K
1	5K	8K	11K	24K
2	7K	10K	13K	30K
$P(y=y_j)$	15K	24K	33K	72K

Given:

We know that,

In Joint probability Mass Function,

$$\sum_{i=0}^{2}\sum_{j=1}^{3} p(x_i, y_j) = 1$$

$72K = 1$

$K = 1/72$

x \ y	1	2	3	$P(X=x_i)$
0	$3/72$	$6/72$	$9/72$	$18/72$
1	$5/72$	$8/72$	$1/72$	$24/72$
2	$7/72$	$10/72$	$13/72$	$30/72$
$P(Y=y_j)$	$15/72$	$24/72$	$33/72$	1

i) Marginal Distribution of x:

$P(x=0) = 18/72$; $P(x=1) = 24/72$; $P(x=2) = 30/72$.

Marginal Distribution of y:

$P(y=1) = 15/72$; $P(y=2) = 24/72$; $P(y=3) = 33/72$

ii) Probability Distribution of x+y:

x+y	Probability
1 (P_{01})	$3/72$
2 $(P_{02}+P_{11})$	$6/72 + 7/72 = 11/72$
3 $(P_{03}+P_{12}+P_{21})$	$9/72 + 8/72 = 24/72$
4 $(P_{13}+P_{22})$	$11/72 + 10/72 = 21/72$
5 (P_{23})	$13/72$
Total =	1

iii) $P(x+y>3) = P(x+y=4) + P(x+y=5)$

$= 21/72 + 13/72$

$= 34/72 = P(x+y>3)$

(or)

$P(x+y>3) = P_{13} + P_{22} + P_{23}$

$= 34/72$ from table,

iv) conditional Distribution x on y:

$$P(x=0/y=1) = \frac{P_{01}}{P(y=1)} = \frac{3/72}{15/72} = 3/15 = 1/5$$

$$P(x=1/y=1) = \frac{P_{11}}{P(y=1)} = \frac{5/72}{15/72} = 5/15 = 1/3$$

$$P(x=2/y=1) = \frac{P_{21}}{P(y=1)} = \frac{7/72}{15/72} = 7/15$$

$$P(x=0/y=2) = \frac{P_{02}}{P(y=2)} = \frac{6/72}{24/72} = 6/24 = 1/6$$

$$P(x=1/y=2) = \frac{P_{12}}{P(y=2)} = \frac{8/72}{24/72} = 8/24 = 1/3$$

$$\Rightarrow P(x=2/y=2) = \frac{P_{22}}{P(y=2)} = \frac{\frac{10}{72}}{\frac{24}{72}} = \frac{5}{12}$$

$$P(x=0/y=3) = \frac{P_{03}}{P(y=3)} = \frac{9/72}{33/72} = 9/33$$

$$P(x=1/y=3) = \frac{P_{13}}{P(y=3)} = \frac{11/72}{33/72} = 11/33$$

$$P(x=2/y=3) = \frac{P_{23}}{P(y=3)} = \frac{13/72}{33/72} = 13/33$$

<u>Conditional</u> <u>Distribution</u> <u>of</u> <u>y</u> <u>on</u> <u>x</u>:

$$P(y=1/x=0) = \frac{P_{01}}{P(x=0)} = \frac{3/72}{18/72} = 3/18 = 1/6$$

$$P(y=2/x=0) = \frac{P_{02}}{P(x=0)} = \frac{6/72}{18/72} = 6/18 = 1/3$$

$$P(y=3/x=0) = \frac{P_{03}}{P(x=0)} = \frac{9/72}{18/72} = 9/18 = 1/2$$

$$P(y=1/x=1) = \frac{P_{11}}{P(x=1)} = \frac{5/72}{24/72} = 5/24$$

$$P(y=2/x=1) = \frac{P_{12}}{P(x=1)} = \frac{8/72}{24/72} = 8/24 = 1/3$$

$$P(y=3/x=1) = \frac{P_{13}}{P(x=1)} = \frac{11/72}{24/72} = 11/24$$

$$P(y=1/x=2) = \frac{P_{21}}{P(x=1)} = \frac{7/72}{30/72} = 7/30$$

$$P(Y=2/X=2) = \frac{P_{22}}{P(X=2)} = \frac{10/72}{30/72} = 10/30 = 1/3$$

$$P(Y=3/X=2) = \frac{P_{23}}{P(X=2)} = \frac{13/72}{30/72} = 13/30$$

3. A Joint Distribution of $f(x,y) = \frac{x+y}{21}$, $x = 1, 2, 3$; $y = 1, 2$.

i) Find Marginal Distribution.

ii) Find $E(x, y)$.

Sol

X \ Y	1	2	$P(X=x_i)$
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
$P(Y=y_j)$	9/21	12/21	1

i) Marginal Distribution of X:

$P(X=1) = 5/21$; $P(X=2) = 7/21$; $P(X=3) = 9/21$.

Marginal Distribution of Y:

$P(Y=1) = 9/21$; $P(Y=2) = 12/21$

ii) $E(xy) = \sum_{i=1}^{3} \sum_{j=1}^{2} xy\, P(x_i, y_j)$

$= (1 \times 1 \times 2/21) + (1 \times 2 \times 3/21) + (2 \times 1 \times 3/21) + (2 \times 2 \times 4/21) + (3 \times 1 \times 4/21) + (3 \times 2 \times 5/21)$

$= 2/21 + 6/21 + 6/21 + 16/21 + 12/21 + 30/21$

$= \frac{30+28+14}{21}$

$= \frac{72}{21}$

4. A joint p.d.f of random variable $f(x,y)$ is given by,

$f(x,y)$ is given by,

$f(x,y) = kxye^{-(x^2+y^2)}$, $x \geq 0$, $y \geq 0$

i) find k.

ii) Prove that x and y are independent.

Solution:

Given:

$f(x,y)$ is joint p.d.f

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

$$= \int_{0}^{\infty}\int_{0}^{\infty} kxye^{-(x^2+y^2)}\,dx\,dy = 1$$

$$= k\int_{0}^{\infty}\int_{0}^{\infty} xye^{-x^2}e^{-y^2}\,dx\,dy = 1$$

$$= k\int_{0}^{\infty} xe^{-x^2}dx \int_{0}^{\infty} ye^{-y^2}dy = 1 \rightarrow (1)$$

To find, $\int_{0}^{\infty} xe^{-x^2}dx$,

$u = x^2$; $x = 0 \Rightarrow u = 0$

$du/dx = 2x$; $x = \infty \Rightarrow u = \infty$

$du/2 = x\,dx$

$$= \int_{0}^{\infty} e^{-u}\,du/2 = 1/2\left(\frac{e^{-u}}{-1}\right)_{0}^{\infty}$$

$$= -1/2(0-1)$$

$$= 1/2$$

$$\therefore \int_{0}^{\infty} xe^{-x^2}dx = 1/2$$

From ①,

$K(1/2)(1/2) = 1$

$K/4 = 1$

$\boxed{K = 4}$

$\therefore f(x,y) = 4xye^{-(x^2+y^2)}, \; x \geq 0, y \geq 0$

ii) $f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$= \int_{0}^{\infty} 4xye^{-(x^2+y^2)}\,dy$

$= 4xe^{-x^2}\int_{0}^{\infty} ye^{-y^2}\,dy$

$= 4xe^{-x^2}(1/2)$

$= 2xe^{-x^2}$

$f(x) = 2xe^{-x^2}$

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

$= \int_{0}^{\infty} 4xye^{-x^2-y^2}\,dx$

$= 4ye^{-y^2}\int_{0}^{\infty} xe^{-x^2}\,dx$

$= 4ye^{-y^2}(1/2)$

$f(y) = 2ye^{-y^2}$

$f(x).f(y) = (2xe^{-x^2})(2ye^{-y^2})$

$= 4xye^{-x^2-y^2}$

$= 4xye^{-(x^2+y^2)}$

$f(x).f(y) = f(x,y)$

$\therefore$ x and y are independent

5. A Joint p.d.f of Random variable $f(x,y)$ is $f(x,y) = e^{-(x+y)}$, $0<x, y<\infty$. Prove that x and y are independent.

Solution:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{0}^{\infty} e^{-x-y}\,dy$$

$$= e^{-x}\int_{0}^{\infty} e^{-y}\,dy$$

$$= e^{-x}\left(\frac{e^{-y}}{-1}\right)_0^{\infty}$$

$$= e^{-x}(0+1)$$

$$f(x) = e^{-x}$$

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

$$= \int_{0}^{\infty} e^{-x}\cdot e^{-y}\,dx$$

$$= e^{-y}\left(\frac{e^{-x}}{-1}\right)_0^{\infty}$$

$$= e^{-y}(0+1)$$

$$f(y) = e^{-y}$$

$$f(x).f(y) = e^{-x}.e^{-y}$$

$$= e^{-x-y}$$

$$= e^{-(x+y)} = f(x,y)$$

$$f(x).\,f(y) = f(x,y)$$

$\therefore$ x and y are independent.

Hence proved.

x and y are independent.

6. If the joint p.d.f of two dimensional random variable,

$$f(x,y) = \begin{cases} x^2 + xy/3 \ ; & 0<x<1,\ 0<y<2 \\ 0 \ ; & \text{Otherwise} \end{cases}$$

i) Find $P(x > 1/2)$

ii) Find $P(x>1)$

iii) Find $P(y < x)$

iv) Find $P(y<1/2 \,/\, x < 1/2)$

v) Find $P(x+y \geq 1)$

vi) Find the conditional density function.

Solution:

Marginal Distribution of x:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^2 (x^2 + xy/3)\,dy$$

$$= \left(x^2 y + xy^2/6\right)_0^2$$

$$= 2x^2 + 2x/3$$

$$f(x) = 2x^2 + 2x/3$$

Marginal density Function of y:

$$f(y) = \int_0^1 (x^2 + xy/3)\,dx$$

$$= \left(x^3/3 + x^2 y/6\right)_0^1$$

$$= 1/3 + y/6$$

$$= \frac{2+y}{6}$$

$$f(y) = \frac{2+y}{6}$$

i) $P(x>1/2) = \int_{1/2}^{1}\left(2x^2+2x/3\right)dx$

$= \left(\frac{2x^3}{3}+\frac{x^2}{3}\right)_{1/2}^{1}$

$= \left(2/3+1/3\right)-\left(\frac{1/4}{3}+1/12\right)$

$= 2/3+1/3-1/12-1/12$

$= 1-1/6$

$= 5/6$

ii) $P(y>1) = \int_{1}^{2}\left(\frac{2+y}{6}\right)dy$

$= 1/6\left(2y+y^2/2\right)_1^2$

$= 1/6\left(4+2-2-1/2\right)$

$= 1/6\left(7/2\right)$

$= 7/12$

iii) $P(y<x) = \int_0^1\int_0^x\left(x^2+\frac{xy}{3}\right)dy\,dx$

$= \int_0^1\left(x^2y+\frac{xy^2}{6}\right)_0^x dx$

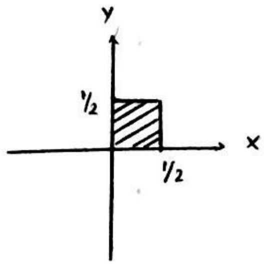
$= \int_0^1\left(x^3+x^3/6\right)dx$

$= \int_0^1 7x^3/6\,dx$

$= \left(\frac{7x^4}{24}\right)_0^1$

$= 7/24$

iv) $P(y<1/2 \,/\, x<1/2) = \dfrac{P(x<1/2 \cap y<1/2)}{P(x<1/2)} \rightarrow (1)$



$$= \int_0^{1/2}\int_0^{1/2} \left(x^2 + xy/3\right) dx\,dy$$

$$= \int_0^{1/2} \left(x^3/3 + x^2y/6\right)_0^{1/2} dy$$

$$= \int_0^{1/2} \left(1/24 + y/24\right) dy$$

$$= \left(y/24 + y^2/48\right)_0^{1/2}$$

$$= 1/48 + 1/192$$

$$= 5/192$$

$P(x<1/2) = 1 - P(x \geq 1/2)$

$$= 1 - 5/6$$

$$= 1/6$$

$\therefore$ from (1),

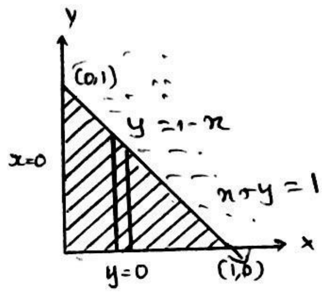
$\Rightarrow \dfrac{5/192}{1/6} = 5/192 \times 6/1 = 30/192$

$$= 15/96$$

$P(y<1/2 \,/\, x<1/2) = 5/32$

v) $P(x+y \geq 1)$

$= 1 - P(x+y<1) \rightarrow (2)$



$$P(x+y<1) = \int_0^1\int_0^{1-x} \left(x^2 + xy/3\right) dy\,dx$$

$$= \int_0^1 \left(x^2y + xy^2/6\right)_0^{1-x} dx$$

$$= \int_0^1 \left(x^2 - x^3 + \frac{x + x^3 - 2x^2}{6} \right) dx$$

$$= \int_0^1 \left(x^2 - x^3 + x/6 + x^3/6 - 2x^2 \right)$$

$$= \int_0^1 \left(-x^2 - 5x^3/6 + x/6 \right) dx$$

$$= \int_0^1 \left(\frac{-5x^3}{6} - x^2 + x/6 \right) dx$$

$$= \left(\frac{-5x^4}{24} - x^3/3 + x^2/12 \right)_0^1$$

$$= -5/24 - 1/3 + 1/12$$

$$= -5/24 - 3/12$$

$$= 7/72$$

$$\therefore \quad P(x+y \geq 1) = 1 - P(x+y < 1)$$

$$= 1 - 7/72$$

$$= 65/72$$

vi) Conditional Density function of x:

$$f(x/y) = \frac{f(x,y)}{f(y)} = \frac{x^2 + xy/3}{\frac{2+y}{6}}$$

$$= \frac{(3x^2 + xy)2}{2+y}$$

$$= \frac{6x^2 + 2xy}{2+y}$$

$$f(y/x) = \frac{f(x,y)}{f(x)} = \frac{x^2 + xy/3}{2x^2 + 2x/3} = \frac{3x^2 + xy}{6x^2 + 2x}$$

7. If x and y are two dimensional random variable having p.d.f

$$f(x,y) = \begin{cases} \frac{1}{8}(6-x-y) & ; \ 0<x<2, \ 2<y<4 \\ 0 & ; \ \text{otherwise} \end{cases}$$

i) Find $P((x<1)\cap(y<3))$

ii) Find $P(x<1 / y<3)$

iii) Find $P((x+y)<3)$

Solution:

i) $P((x<1)\cap(y<3)) = \int_0^1\int_2^3 \frac{1}{8}(6-x-y)\,dy\,dx$

$= \frac{1}{8}\int_0^1 \left(6y-xy-\frac{y^2}{2}\right)_2^3 dx$

$= \frac{1}{8}\int_0^1 (18-3x-\frac{9}{2}-12+2x+2)\,dx$

$= \frac{1}{8}\int_0^1 (8-x-\frac{9}{2})\,dx$

$= \frac{1}{8}\int_0^1 (\frac{7}{2}-x)\,dx$

$= \frac{1}{8}\left(\frac{7x}{2}-\frac{x^2}{2}\right)_0^1$

$= \frac{1}{8}\left(\frac{7}{2}-\frac{1}{2}\right)$

$= \frac{1}{8}\left(\frac{6}{2}\right)$

$= \frac{3}{8}$

ii) $p(x<1 / y<3) = \frac{P(x<1)\cap(y<3)}{P(y<3)} \rightarrow (1)$

$P(y<3) = \int_2^3 f(y)\,dy$

Marginal Density of Y,

$$f(y) = \int_0^2 \frac{1}{8}(6-x-y)\,dx$$

$$= \frac{1}{8}\left(6x - \frac{x^2}{2} - xy\right)_0^2$$

$$= \frac{1}{8}(12-2-2y)$$

$$= \frac{1}{8}(10-2y)$$

$$= \frac{10-2y}{8}$$

$$P(y<3) = \frac{1}{8}\int_2^3 (10-2y)\,dy$$

$$= \frac{1}{8}\left(10y - \frac{2y^2}{2}\right)_2^3$$

$$= \frac{1}{8}(30-9-20+4)$$

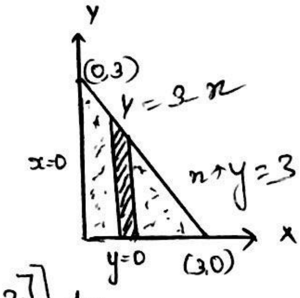
$$= \frac{1}{8}(10-5)$$

$$= \frac{5}{8}$$

$\therefore$ from ①, $\frac{3/8}{5/8} = 3/5 = P\left[X<1 / y<3\right]$

iii) $P(x+y<3)$

$$= \int_0^2 \int_2^{3-x} \frac{1}{8}(6-x-y)\,dy\,dx$$



$$= \frac{1}{8}\int_0^2 \left(6y - xy - \frac{y^2}{2}\right)_2^{3-x} dx = \frac{1}{8}\int_0^2 \left[\left[6(3-x) - x(3-x) - \frac{(3-x)^2}{2}\right] - \left[12-2x-2\right]\right] dx$$

$$= \frac{1}{8}\int_0^2 \left[18 - 6x - \cancel{3x} + x^2 - \frac{9}{2} - \frac{x^2}{2} + \cancel{\frac{6x}{2}} + 2x - 10\right] dx.$$

$$= \frac{1}{8}\int_0^2 \left[\frac{x^2}{2} - 4x + \frac{7}{2}\right] dx$$

$$= \frac{1}{8}\left[\frac{x^3}{6} - 4\frac{x^2}{2} + \frac{7}{2}x\right]_0^2$$

$$= \frac{1}{8}\left[\frac{8}{6} - \frac{16}{2} + \frac{14}{2}\right]$$

$$= \frac{1}{8}\left[\frac{4}{3} - 8 + 7\right]$$

$$= \frac{1}{8}\left[\frac{4}{3} - 1\right]$$

$$= \frac{1}{8}\left[\frac{4-3}{3}\right]$$

$$= \frac{1}{8}\left(\frac{1}{3}\right)$$

$$= \frac{1}{24}$$

8. A joint p.d.f $f(x,y) = \begin{cases} 8xy; & 0<x<y<1 \\ 0; & \text{otherwise} \end{cases}$

i) Find Marginal, p.d.f of x & y

ii) Prove that, x and y are independent.

<u>Solution:</u>

Marginal p.d.f of x,

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_{x}^{1} 8xy\,dy$$

$$= \left(\frac{8xy^2}{2}\right)_x^1$$

$$= \left(\frac{8x}{2} - \frac{8x^3}{2}\right)$$

$= 4x - 4x^3$

Marginal p.d.f of Y,

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

$$= \int_0^y 8xy\,dx$$

$$= \left(\frac{8x^2y}{2}\right)_0^y$$

$$= (4x^2y)_0^y$$

$$= 4y^3$$

$f(x).f(y) = (4x-4x^3).(4y^3) \neq 8xy \neq f(x,y)$

$\therefore$ x and y are not independent.

---

9. Given $f(x,y) = \begin{cases} cx(x-y) \ ; & 0<x<2 \\ & -x<y<x \\ 0 & ; \text{Otherwise} \end{cases}$ fixed.

i) Evaluate c.

ii) Find $f_x(x)$

iii) Find $F(y/x)$

iv) Find $f_y(y)$

Solution:

i) since $f(x,y)$ is joint p.d.f

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

$$\Rightarrow \int_0^2\int_{-x}^{x} cx(x-y)\,dy\,dx = 1$$

$$\Rightarrow c\int_0^2\int_{-x}^{x} (x^2-xy)\,dy\,dx = 1$$

$$\Rightarrow c\int_0^2 \left[x^2y - xy^2/2\right]_{-x}^{x} dx = 1$$

$$\Rightarrow c\int_0^2 \left(x^3 - x^3/2\right) - \left(-x^3 - x^3/2\right)$$

$$\Rightarrow c\int_0^2 2x^3 dx = 1$$

$$\Rightarrow c\left(\frac{2x^4}{4}\right)_0^2 = 1$$

$$\Rightarrow c\left(x^4/2\right)_0^2 = 1$$

$$\Rightarrow c\left(16/2\right) = 1$$

$$8c = 1$$

$$c = 1/8$$

ii) $f_x(x) = \int_{-x}^{x} f(x,y)\,dy$

$$= 1/8 \int_{-x}^{x} (x^2 - xy)\,dy$$

$$= 1/8 \left(x^2 y - xy^2/2\right)_{-x}^{x}$$

$$= 1/8 \left(x^3 - \cancel{x^3/2} + x^3 + \cancel{x^3/2}\right)$$

$$= 1/8 (2x^3)$$

$$= x^3/4$$

iv) $f_y(y) = \int_0^2 f(x,y)\,dx$

$$= 1/8 \int_0^2 (x^2 - xy)\,dx$$

$$= 1/8 \left(x^3/3 - x^2 y/2\right)_0^2$$

$$= \frac{1}{8}\left(\frac{8}{3} - 2y\right)$$

$$= \frac{1}{8}\left(\frac{8-6y}{3}\right)$$

$$= \frac{1}{24}(8-6y)$$

$$= \frac{1}{4}\left(\frac{4}{3} - y\right)$$

$$= \frac{1}{12}(4-3y)$$

iii) $$F(y/x) = \frac{f(x,y)}{f(x)}$$

$$= \frac{\frac{1}{8}(x^2 - xy)}{x^3/4}$$

$$= \left(\frac{x^2}{8} - \frac{xy}{8}\right)\left(\frac{4}{x^3}\right)$$

$$= \left(\frac{x(x-y)}{8}\right)\left(\frac{4}{x^3}\right)$$

$$F(y/x) = \frac{x-y}{2x^2}$$

CO-VARIANCE:

If x and y are two Random Variables, the co-variance between them,

$$COV(X,Y) = E(XY) - E(X)E(Y)$$

Note:

If x and y are independent, then

$$E(XY) = E(X)E(Y)$$

$$\Rightarrow \quad COV(XY) = 0$$

---

1. If x has Mean = 4, variance = 9, while y Mean has = −2, variance = 5. and the two are independent.

Find i) $E(XY)$ ii) $E(XY^2)$.

Solution:

i) Mean $E(x) = 4$, Mean $E(y) = -2$

$Var(x) = 9$, $Var(y) = 5$

Since x and y are independent.

$E(xy) = E(x).E(y)$

$E(xy) = 4(-2)$

$= -8$

ii) We know that, $var(y) = E(y^2) - [E(y)]^2$

$5 = E(y^2) - 4$

$5+4 = E(y^2)$

$E(y^2) = 9$

$E(xy^2) = E(x).E(y^2)$

$E(xy^2) = 4 \times 9$

$= 36$

---

2. Let x_1 and x_2 has joint p.m.f $P(x_1, x_2) = \frac{x_1+2x_2}{18}$ where $x_1 x_2 = 1, 2$. Find $cov(x_1 x_2)$.

Solution:

Given:

$P(x_1, x_2) = \frac{x_1+2x_2}{18}$, $x_1 x_2 = 1, 2$

x_1 \ x_2	1	2	$P(x_1 = x_1)$
1	3/18	5/18	8/18
2	4/18	6/18	10/18
$P(y_2 = y_2)$	7/18	11/18	1
	$P(x_2=1)$	$P(x_2=2)$	

$E(x_1) = \sum x_1 p(x_1)$

$= (1) 8/18 + 2 \times 10/18$

$= 28/18$

$E(X_2) = \sum x_2 p(x_j)$

$= (1)\left(\frac{7}{18}\right) + 2\left(\frac{11}{18}\right)$

$= \frac{7}{18} + \frac{22}{18}$

$= \frac{29}{18}$

$E(x_1, x_2) = \left(1 \times \frac{3}{18}\right) + \left(2 \times \frac{5}{18}\right) + \left(2 \times \frac{4}{18}\right) + \left(4 \times \frac{6}{18}\right)$

$= \frac{3}{18} + \frac{10}{18} + \frac{8}{18} + \frac{24}{18}$

$= \frac{45}{18}$

$COV(x_1, x_2) = E(X_1 X_2) - E(X_1)\,E(X_2)$

$= \frac{45}{18} - \left(\frac{28}{18} \cdot \frac{29}{18}\right)$

$= \frac{45}{18} - \frac{812}{324}$

$= \frac{-2}{324}$

$COV(x_1, x_2) = -0.016$

CORRELATION:

If the change in one variable affects the change in other variable then the variables are said to be correlated.

TYPES:

- positive and Negative correlation.
- simple, partial correlation
- Linear, Non-Linear correlation.

Karl Pearson co.efficient:

$$\text{Correlation co.efficient } \gamma = \frac{cov(x,y)}{\sqrt{var(x)} \cdot \sqrt{var(y)}}$$

$$= \frac{E(xy) - E(x)E(y)}{\sigma_x \cdot \sigma_y}$$

where $E(x) = \frac{\Sigma x}{n} = \bar{x}$,

$E(y) = \frac{\Sigma y}{n} = \bar{y}$,

$E(xy) = \frac{\Sigma xy}{n}$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2}$

$\sigma_y = \sqrt{\frac{1}{n}\Sigma y^2 - \bar{y}^2}$

Note:

i) correlation coefficient always lies between -1 to +1, $-1 \le r \le 1$

ii) $r = -1 \Rightarrow$ Perfect Negative correlation.

$r = 1 \Rightarrow$ Perfect positive correlation.

$r = 0 \Rightarrow$ No correlation (uncorrelated).

1. calculate the correlation coefficient, for the following heights in inches of father (x) and son (y).

x	65	66	67	68	69	70	72
y	67	68	65	72	72	69	71.

soln:

x	y	xy	x^2	y^2
65	67	4355	4325	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4950	4761
72	71	5112	5184	5041
477	484	33004	32539	33508

$\bar{x} = E(x) = \frac{\Sigma x}{n} = \frac{477}{7}$

$= 68.14$

$\bar{y} = E(y) = \frac{\Sigma y}{n} = \frac{484}{7}$

$= 69.14$

$E(xy) = \frac{\Sigma xy}{n} = \frac{33004}{7}$

$= 4714.85$

$cov(xy) = E(xy) - E(x)E(y)$

$= 4714.85 - (168.14)(169.14)$

$= 3.65$

$\sigma_x = \sqrt{1/n \, \Sigma x^2 - \bar{x}^2}$

$= \sqrt{\frac{31539}{7} - 168.14}$

$= 2.25$

$\sigma_y = \sqrt{1/n \, \Sigma y^2 - \bar{y}^2}$

$= \sqrt{\frac{33508}{7} - (69.14)^2}$

$= 2.55$

Correlation coefficient $r = \frac{cov(x,y)}{\sigma_x \sigma_y}$

$= \frac{3.65}{(2.25)(2.55)}$

$r = 0.636.$

2. Two Dimensional random variable x, y have joint p.d.f,

$$f(x,y) = \begin{cases} 2-x-y \ ; & 0<x<1 \\ & 0<y<1 \\ 0 & ; \text{Otherwise} \end{cases}$$

Find γ coefficient of correlation.

Solution:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^1 (2-x-y)\,dy$$

$$= \left(2y - xy - \frac{y^2}{2}\right)_0^1$$

$$= 2-x-\frac{1}{2}$$

$$= \left(\frac{4-2x-1}{2}\right)$$

$$f(x) = \frac{3}{2} - x$$

$$f(y) = \int_0^1 (2-x-y)\,dx$$

$$= \left(2-y-\frac{1}{2}\right)$$

$$= \frac{3}{2} - y$$

$$E(x) = \int_0^1 x \cdot \left(\frac{3}{2} - x\right) dx$$

$$= \int_0^1 \frac{3}{2}x - x^2\,dx$$

$$= \left(\frac{3x^2}{4} - \frac{x^3}{3}\right)_0^1$$

$$= \frac{3}{4} - \frac{1}{3}$$

$$= \frac{9-4}{12}$$

$E(x) = 5/12$. ||| ly $E(y) = 5/12$

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

$$= \int_0^1 \left(\tfrac{3}{2}x^2 - x^3\right) dx$$

$$= \left(\tfrac{\cancel{3}}{2} \cdot \tfrac{x^3}{\cancel{3}} \cdot \tfrac{x^4}{4}\right)_0^1$$

$$= \tfrac{1}{2} - \tfrac{1}{4}$$

$$= \tfrac{1}{4}$$

$\therefore$ ${}^{\text{lly}}$, $E(y^2) = 1/4$

$$\text{var}(x) = E(x^2) - (E(x))^2$$

$$= \tfrac{1}{4} - \tfrac{25}{144}$$

$$= \tfrac{11}{144}.$$

$$\text{var}(y) = E(y^2) - (E(y))^2$$

$$= \tfrac{1}{4} - \tfrac{25}{144}$$

$$= \tfrac{11}{144}$$

$$E(xy) = \int_0^1 \int_0^1 xy(2-x-y)\,dx\,dy$$

$$= \int_0^1 \int_0^1 (2xy - x^2y - xy^2)\,dx\,dy$$

$$= \int_0^1 \left(\frac{2x^2y}{2} - \frac{x^3y}{3} - \frac{x^2y^2}{2}\right) dy$$

$$= \int_0^1 \left(y - y/3 - y^2/2\right) dy$$

$$= \left(y^2/2 - y^2/6 - y^3/6\right)_0^1$$

$$= 1/6.$$

$cov(xy) = E(xy) - E(x)E(y)$

$= 1/6 - (5/12)(5/12)$

$= 1/6 - 25/144$

$= \frac{24-25}{144}$

$= -1/144$

correlation coefficient γ

$$\gamma = \frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$$

$$= \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}}$$

$$= \frac{-1/144}{11/144}$$

$\gamma = -1/11.$

---

3. A Random variable x, y has a joint p.d.f $f(x,y) = \begin{cases} x+y; & 0<x<1 \\ & 0<y<1 \\ 0; & \text{otherwise} \end{cases}$ Find γ

solution:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$

$$= \int_0^1 (x+y)\,dy$$

$$= \left(xy + y^2/2\right)_0^1$$

$$= (x + 1/2)$$

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

$$= \int_0^1 (x+y)\,dx$$

$$= \left(\frac{x^2}{2} + xy\right)_0^1$$

$$= \frac{1}{2} + y$$

$$E(X) = \int_0^1 x\left(x + \frac{1}{2}\right)dx$$

$$= \int_0^1 \left(x^2 + \frac{x}{2}\right)dx$$

$$= \left(\frac{x^3}{3} + \frac{x^2}{4}\right)_0^1$$

$$= \frac{1}{3} + \frac{1}{4}$$

$$= \frac{7}{12}$$

Similarly, $E(Y) = 7/12$

$$E(X^2) = \int_0^1 x^2\left(x + \frac{1}{2}\right)dx$$

$$= \int_0^1 \left(x^3 + \frac{x^2}{2}\right)dx$$

$$= \left(\frac{x^4}{4} + \frac{x^3}{6}\right)_0^1$$

$$= \frac{1}{4} + \frac{1}{6}$$

$$= \frac{5}{12}$$

Similarly $E(Y^2) = 5/12$

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

$$= \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$$

Similarly $var(y) = 11/144$

$$E(xy) = \int_0^1 \int_0^1 xy(x+y)\,dx\,dy$$

$$= \int_0^1 \int_0^1 (x^2y + xy^2)\,dx\,dy$$

$$= \int_0^1 \left(\frac{x^3y}{3} + \frac{x^2y^2}{2}\right)_0^1 dy$$

$$= \int_0^1 \left(y/3 + y^2/2\right) dy$$

$$= \left(y^2/6 + y^3/6\right)_0^1$$

$$= 1/6 + 1/6$$

$$= 1/3$$

$$Cov(x,y) = E(xy) - E(x)E(y)$$

$$= 1/3 - (7/12)(7/12)$$

$$= 1/3 - 49/144$$

$$= -1/144$$

Correlation coefficient $r = \dfrac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}$

$$= \frac{-1/144}{\sqrt{11/144}\,.\,\sqrt{11/144}}$$

$$= \frac{-1/144}{11/144}$$

$$r = -1/11$$

4. $f(x,y) = \begin{cases} \frac{1}{8}(6-x-y) ; & 0<x<2 \\ & 2<y<5 \\ 0 ; & \text{Otherwise} \end{cases}$ Find the correlation between x & y.

Solution:

$$f(x) = \int_2^5 \frac{1}{8}(6-x-y)\,dy$$

$$= \frac{1}{8}\int_2^5 (6-x-y)\,dy$$

$$= \frac{1}{8}\left(6y - xy - \frac{y^2}{2}\right)_2^5$$

$$= \frac{1}{8}\left(30 - 5x - \frac{25}{2} - 12 + 2x + 2\right)$$

$$= \frac{1}{8}\left(-3x + 20 - \frac{25}{2}\right)$$

$$= \frac{1}{8}\left(-3x + \frac{40-25}{2}\right)$$

$$= \frac{1}{8}\left(-3x + \frac{40-25}{2}\right)$$

$$= \frac{1}{8}\left(-3x + \frac{15}{2}\right)$$

$$= \frac{3}{8}\left(-x + \frac{5}{2}\right)$$

$$= \frac{3}{8}\left(\frac{5}{2} - x\right)$$

$$= \frac{3}{8}\left(\frac{5-2x}{2}\right)$$

$$f(x) = \frac{3}{16}(5-2x)$$

$$f(y) = \int_0^2 \frac{1}{8}(6-x-y)\,dx$$

$$= \frac{1}{8}\int_0^2 [6-x-y]\,dx$$

$$= \frac{1}{8}\left(6x - \frac{x^2}{2} - xy\right)_0^2$$

$$= \frac{1}{8}(12-2-2y)$$

$= \frac{1}{8}(10-2y)$

$= \frac{2}{8}(5-y)$

$= \frac{1}{4}(5-y)$

$E(x) = \frac{3}{8}\int_0^2 (6x - x^2 - xy)\left(\frac{5}{2}x - x^2\right)dx$

$= \frac{3}{8}\left(\frac{5x^2}{4} - \frac{x^3}{3}\right)_0^2$

$= \frac{3}{8}\left(5 - 8/3\right)$

$= \frac{\cancel{3}}{8}\left(7/\cancel{3}\right)$

$= 7/8$

$E(x^2) = \frac{3}{8}\int_0^2 \left(\frac{5}{2}x^2 - x^3\right)dx$

$= \frac{3}{8}\left(\frac{5x^3}{6} - \frac{x^4}{4}\right)_0^2$

$= \frac{3}{8}\left(40/6 - 16/4\right)$

$= \frac{3}{8}\left(\frac{40(2) - 16(3)}{12}\right)$

$= \frac{3}{8}\left(\frac{80-48}{12}\right)$

$= \frac{3}{8}\left(32/12\right)$

$= \frac{3}{8}\left(\overset{4}{\underset{1}{\cancel{32}}}/12\right)$

$= 12/12$

$E(x^2) = 1$

$$E(Y) = \frac{3}{8}\int_2^5 \left(\frac{5}{4} - \frac{y}{4}\right) dy$$

$$= \frac{3}{8}\left(\frac{5y}{4} - \frac{y^2}{8}\right)_2^5$$

$$= \frac{3}{8}\left(\frac{25}{4} - \frac{25}{8} - \frac{10}{4} + \frac{4}{8}\right)$$

$$= \frac{3}{8}\left(\frac{25}{4} - \frac{25}{8} - \frac{5}{2} + \frac{1}{2}\right)$$

$$= \frac{3}{8}\left(\frac{40-25}{8} + \frac{1}{2}\right)$$

$$E(Y^2) = \int_2^5 \left(\frac{5y^2}{4} - y^3\right) dy$$

$$= \left(\frac{5y^3}{12} - \frac{y^4}{4}\right)_2^5$$

$$= \left(\frac{5\times125}{12} - \frac{625}{4} - \frac{40}{12} + \frac{16}{4}\right)$$

$$= \left(\frac{625}{12} - \frac{625}{4} - \frac{40}{12} + 4\right)$$

$$= \left(\frac{-1250}{12} - \frac{10}{3} + 4\right)$$

$$= \frac{-1210}{12} + 4$$

$$= \frac{-1162}{12}$$

$$E(Y^2) = 10.6$$

$$E(XY) = \int_0^2\int_2^5 xy\, \frac{1}{8}\,(6-x-y)\, dx\, dy$$

$$= \frac{1}{8}\int_0^2\int_0^5 (6xy - x^2y - xy^2)\, dx\, dy$$

$$= 1/8 \int_0^2 \left(\frac{6x^2y}{2} - \frac{x^3y}{2} - \frac{x^2y^2}{2} \right)_0^5 dy$$

$$= 1/8 \int_0^2 \left(\frac{150y}{2} - \frac{125y}{3} - \frac{25y^2}{2} \right) dy$$

$$= 1/8 \left(\frac{150y^2}{4} - \frac{125y^2}{6} - \frac{25y^3}{6} \right)_0^2$$

$$= 1/8 \left(150 - \frac{125 \times 2}{3} - \frac{100}{3} \right)$$

$$= 1/8 \left(150 - 250/3 - 100/3 \right)$$

$$= 1/8 \left(150 - 150/3 \right)$$

$$= 1/8 (150 - 50)$$

$$= 1/8 (100)$$

$$= 50/\cancel{8}$$

$$= 25/4$$

$$= 5/2$$

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

$$= 1 - 49/64$$

$$= \frac{64-49}{64}$$

$$= 15/64$$

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

$$= 10.6 - \frac{729}{64}$$

$$= 10.6 - 11.39$$

$$Var(y) = -0.79$$

$$Cov(x,y) = E(xy) - E(x)E(y)$$

$= 5/2 - (7/8)(27/8)$

$= 5/2 - 189/64$

$= \frac{160-189}{64}$

$= -29/64$

$= -0.45$

Correlation Coefficient $r = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x)}\sqrt{\text{var}(y)}}$

$= \frac{-0.45}{\sqrt{0.234}\sqrt{0.234}}$

$r = 0.03$

---

5. Let x and y are discrete Random variable with probability function. $f(x) = \frac{x+y}{21}$, $x = 1,2,3$, $y = 1,3$

i) Find Mean of x & y

ii) Var of x & Var of y

iii) cov(x,y)

iv) R(x,y)

Solution:

x \ y	1	2	$P(x=x_i)$
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
$P(y=y_j)$	9/21	12/21	1

i) Mean of x ⇒ $E(X) = \sum x_i p(x_i)$

$= (1 \times 5/21) + (2 \times 7/21) + (3 \times 9/21)$

$= 5/21 + 14/21 + 27/21$

$= 46/21$

Mean of $y = E(Y) = \sum y_j \, P(y_j)$

$= (1 \times 9/21) + (2 \times 12/21)$

$= 9/21 + 24/21$

$= 33/21$

ii) $E(X^2) = \sum x_i^2 \, P(x_i)$

$= (1 \times 5/21) + (4 \times 7/21) + (9 \times 9/21)$

$= 5/21 + 28/21 + 81/21$

$= 114/21$

$E(Y^2) = \sum y_j^2 \, P(y_j)$

$= (1 \times 9/21) + (4 \times 12/21) +$

$= 9/21 + 48/21$

$= 57/21$

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

$= 114/21 - \frac{1116}{441}$

$= \frac{2394 - 1116}{441}$

$= \frac{1278}{441}$

$Var(X) = 0.63$

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

$= 57/21 - \frac{1089}{441}$

$= \frac{1197 - 1089}{441}$

$= 108/441$

$= 0.24$

$E(xy) = \sum \sum xy\, p(x_i, y_j)$

$= \left(1 \times 1 \times 2/21\right) + \left(1 \times 2 \times 3/21\right) + \left(2 \times 1 \times 3/21\right) + \left(2 \times 2 \times 4/21\right)$
$+ \left(3 \times 1 \times 4/21\right) + \left(3 \times 2 \times 5/21\right)$

$= 2/21 + 6/21 + 6/21 + 16/21 + 12/21 + 30/21$

$= 72/21$

$\mathrm{cov}(x,y) = E(xy) - E(x)E(y)$

$= 72/21 - \left(46/21\right)\left(33/21\right)$

$= -0.013$

iv) $\gamma(x,y) = \frac{\mathrm{cov}(x,y)}{\sqrt{\mathrm{var}(x)}\sqrt{\mathrm{var}(y)}}$

$= \frac{-0.013}{\sqrt{0.04}\sqrt{0.94}}$

$= -0.03$

$\gamma = -0.03$

RANK CORRELATION:

$r = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$ where $d_i = x_i - y_j$
n = no. of terms

1. Calculate Rank correlation for given Data.

x	10	15	12	17	13	16	24	14	22
y	30	42	45	46	33	34	40	35	39

Solution:

x	y	Rank of x	Rank of y	$d_i = x-y$	d_i^2
10	30	9	9	0	0
15	42	5	3	2	4
12	45	8	2	6	36
17	46	3	1	2	4
13	33	7	8	-1	1
16	34	4	7	-3	9
24	40	1	4	-3	9
14	35	6	6	0	0
22	39	2	5	-3	9
					$\Sigma d_i^2 = 72$

Rank Correlation, $\gamma = 1 - \frac{6\Sigma d_i^2}{n(n^2-1)}$

$= 1 - \frac{6(72)}{9(81-1)}$

$= 1 - \frac{432}{720}$

$= 1 - \frac{431}{720}$

$= 1 - 0.6$

$= 0.4$

Rank Correlation $= 0.4$

2. In a beauty contest, ranked by 3 judges in the following order.

Participants	1	2	3	4	5	6	7	8	9	10
Judge A	1	6	5	10	3	2	4	9	7	8
Judge B	3	5	8	4	7	10	2	1	6	9
Judge C	6	4	9	8	1	2	3	10	5	7

Using Rank correlation coefficient. Determine which pair of Judges, have common taste in beauty?

x	y	z	$d_1 = x-y$	$d_2 = y-z$	$d_3 = x-z$	d_1^2	d_2^2	d_3^2
1	3	6	-2	-3	-5	4	9	25
6	5	4	1	1	2	1	1	4
5	8	9	-3	-1	-4	9	1	16
10	4	8	6	-4	2	36	16	4
3	7	1	-4	6	2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	-1	2	1	1	4	1
						$\Sigma d_1^2 = 200$	$\Sigma d_2^2 = 214$	$\Sigma d_3^2 = 60$

Rank correlation between x and y is.

$$r_1(x, y) = 1 - \frac{6 \Sigma d_1^2}{n(n^2-1)}$$

$$= 1 - \left(\frac{6 \times 200}{10(100-1)}\right)$$

$$= 1 - \frac{1200}{990}$$

$$= -210/990$$

$= -7/33$

$= -0.21$

Rank correlation between y and z is

$$r_2(y,z) = 1 - \frac{6\Sigma d_2^2}{n(n^2-1)}$$

$$= 1 - \frac{6(214)}{10(100-1)}$$

$$= -0.29$$

Rank correlation between x and z is

$$r_3(x,z) = 1 - \frac{6\Sigma d_3^2}{n(n^2-1)}$$

$$= 1 - \frac{6(60)}{99}$$

$$= 0.04$$

Since correlation coefficent of 'C and A' is maximum ∴ They have common taste of beauty.

3. Calculate the rank coefficient of correlation for the data.

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70.

Solution:

x	y	Rank of x	Rank of y	$d_i = x-y$	d_i^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25.

80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\Sigma d_i^2 = 72$

Repeated Ranks:

In x term 75 repeated 2 times,

Correlation factor $= \frac{m(m^2-1)}{12}$

$= \frac{2(3)}{12}$

$= 1/2$

$= 0.5$

In x term 64 repeated 3 times,

correlation factor $= \frac{3(9-1)}{12}$

$= 24/12$

$= 2$

In y term 68 repeated 2 times,

correlation factor $= \frac{2(4-1)}{12}$

$= 6/12$

$= 0.5$

Rank correlation $= 1 - \frac{6(\Sigma d^2 + cf)}{n(n^2-1)}$

$= 1 - \frac{6(72+0.5+2+0.5)}{10(99)} = 1 - 6(75)/990 = 540/990$

Rank correlation $= 0.545$

4. Two Independent Random variable x and y defined by,

$f(x) = \begin{cases} 4ax ; & 0 \le x \le 1 \\ 0; & \text{otherwise} \end{cases}$, $f(y) = \begin{cases} 4by ; & 0 \le y \le 1 \\ 0 ; & \text{otherwise} \end{cases}$ show that $u = x+y$ & $v = x-y$ are uncorrelated.

Solution:

Since f(x) is p.d.f

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

$= \int_0^1 4ax\, dx = 1$

$= \left(4a\frac{x^2}{2}\right)_0^1 = 1$

$= 4a/2 = 1$

$= a = 1/2$

$\therefore f(x) = \{2x,\ 0 \le x \le 1\}$

Similarly $f(y) = \{2y,\ 0 \le y \le 1\}$

To prove u and v are uncorrelated.

(i.e) $E(uv) = E(u) \cdot E(v)$

$E(U) = E[x+y] = E(x) + E(y)$

$E(V) = E(x-y) = E(x) - E(y)$

$E(UV) = [E(x) + E(y)][E(x) - E(y)]$

$= E(x^2 - y^2)$

$= E(x^2) - E(y^2)$

$E(x) = \int_0^1 x \cdot 2x\, dx \Rightarrow \int_0^1 2x^2 dx$

$\Rightarrow \left(\frac{2x^3}{3}\right)_0^1 = 2/3$

$$E(x^2) = \int_0^1 x^2 . 2x\,dx = \int_0^1 2x^3\,dx$$

$$= \left(\frac{2x^4}{4}\right)_0^1$$

$$= 1/2$$

$$E(y) = \int_0^1 y.2y\,dy = \int_0^1 2y^2\,dy$$

$$= \left(\frac{2y^3}{3}\right)_0^1$$

$$= 2/3$$

$$E(y^2) = \int_0^1 y^2 2y\,dy = \int_0^1 2y^3\,dy$$

$$= \left(\frac{2y^4}{4}\right)_0^1$$

$$= 1/2$$

$$E(U) = F(x) + E(y) = 4/3$$

$$E(V) = E(x) - E(y) = 0$$

$$E(UV) = 1/4 - 1/4$$

$$= 0$$

$$E(UV) = E(U) . E(V)$$

$\Rightarrow$ U and V are independent

$Cov(u,v) = 0$

$r = 0$

u and v are uncorrelated.

REGRESSION:

Regression is the measure of the average relationship between two or more variables, in terms of original units of data.

Regression line x and y is $(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$

Regression line y on x is.

$$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

$b_{xy} = r\frac{\sigma_x}{\sigma_y}$, $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ are called as Regression co.efficients.

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$$

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

Note:

$$r^2 = b_{xy}b_{yx}$$

$$r = \pm\sqrt{b_{xy}b_{yx}}$$

If θ is the acute angle between 2 regression line,

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

1. From the following data i) Find the two regression, equations. ii) the coefficient of correlation in the marks in economics and statistics iii) the most likely marks in statistics, when marks in economics are 30.

Marks in Economics	25	28	35	32	31	36	29	38	34	32
Marks in statistics	43	46	49	41	36	32	31	30	33	39

Solution:

Given:

x	y	$x - \underset{\bar{x}}{32}$	$y - \underset{\bar{y}}{38}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
25	43	-7	5	-35	49	25
28	46	-4	8	-28	16	64
35	49	3	11	33	9	121
32	41	0	3	0	0	9
31	36	-1	-2	2	1	4
36	32	4	-6	-24	16	36
29	31	-3	-7	21	9	49
38	30	6	-8	-48	36	64
34	33	2	-5	-10	4	25
32	39	0	1	0	0	1
$\Sigma x = 320$	$\Sigma y = 380$	0	0	$\Sigma(x-\bar{x})(y-\bar{y}) = -93$	$\Sigma(x-\bar{x})^2 = 149$	$\Sigma(y-\bar{y})^2 = 398$

$$\bar{x} = \frac{\Sigma x}{n} = 32$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{380}{10} = 38$$

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2} = \frac{-93}{398}$$

$$= -0.23$$

$$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = -93/140$$

$$= -0.66$$

ii) Correlation Coefficient between (economics & statistics)

$$r^2 = b_{xy} \cdot b_{yx}$$

$$r^2 = (-0.23)(-0.66) = 0.152$$

$$r = \pm 0.38$$

i) Regression line x on y :

$(x-\bar{x}) = b_{xy}(y-\bar{y})$

$(x-32) = -0.23(y-38)$

$x = -0.23y + 8.74 + 32$

$x = -0.23y + 40.74$

Regression line y on x :

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$(y-38) = -0.66(x-32)$

$y = -0.66x + 21.12 + 38$

$y = -0.66x + 59.12$

iii) The most likely marks in statistics (y) when marks in economics (x) is 30.

i-e) When $x = 30$.

$y = -0.66 + 59.12$

$= -19.8 + 59.12$

$= 39.32$

---

2. The two lines of Regressions are, $8x - 10y + 66 = 0$, $40x - 18y - 214 = 0$. The variance of x is 6 i) Find Mean values of x and y. ii) correlation coefficient between x and y.

Solution:

Since bt regression lines are passing through the mean value $\bar{x}$ and $\bar{y}$. The point $\bar{x}, \bar{y}$ must satisfy the given two regression line,

$8\bar{x} - 10\bar{y} + 66 = 0 \rightarrow (1) \times 5$

$40\bar{x} - 18\bar{y} - 214 = 0 \rightarrow (2)$

Solve (1) & (2)

$40\bar{x} - 50\bar{y} = -330$

$40\bar{x} - 18\bar{y} = 214$

$-32\bar{y} = -544$

$\bar{y} = 17$.

① in $\bar{y} = 17 \Rightarrow 8\bar{x} - 170 = -66$

$8\bar{x} = 170 - 66$

$\bar{x} = 104/8$

$\bar{x} = 13$

i) Mean of $x = E(x) = 13$

Mean of $y = E(y) = 17$.

① $\rightarrow 8x - 10y + 66 = 0$

$8x = 10y - 66$

$x = 1/8\,(10y - 66)$

② $\rightarrow 40x - 18y - 214 = 0$

$y = 1/18\,(40x - 214)$

y coefficient $\leftarrow b_{xy} = 10/8$

$b_{xy} = 1.25$

x coefficient $\leftarrow b_{yx} = 40/18$

$b_{yx} = 2.22$

$r = \pm\sqrt{b_{xy} b_{yx}}$

$= \pm\sqrt{(1.25)(2.22)}$

$= \pm 1.66$ is not possible

From ②, $8x - 10y + 66 = 0$

$1/10\,(8x + 66) = y$

coefficient of $x = 8/10$

$b_{yx} = 0.8$

From ①, $40x - 18y - 214 = 0$

$1/40\,(18y + 214) = x$

coefficient of $y = 18/40$

$b_{xy} = 0.45$

$r = \pm\sqrt{(0.8)(0.45)}$

$r = \pm\sqrt{0.36}$

$r = 0.6$

3. Find the most likely price in city A, corresponding to the price of Rs. 70 at city B, from the following table.

	city B	City A
Average Price	65	67
Standard deviation σ	2.5	3.5

Solution:

Correlation coefficient is $0.8 = r$

let x denotes the price of city A,

let y denotes the price of city B.

Given:

$\bar{x} = 67$, $\bar{y} = 65$

$\sigma_x = 3.5$ $\sigma_y = 2.5$

$r = 0.8$

Regression line x on y:

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

$$(x-67) = 0.8 \times \frac{3.5}{2.5}(y-65)$$

$$= 0.8 \times 1.4(y-65)$$

$$(x-67) = 1.12y - 72.8$$

$$x = 1.12y - 72.8 + 67$$

$$x = 1.12y - 5.8$$

When $y = 70$

$$x = 1.12 \times 70 - 5.8$$

$$x = 78.40 - 5.8$$

$$x = 72.6$$

TRANSFORMATION OF RANDOM VARIABLE

TRANSFORMATION OF ONE DIMENSIONAL RANDOM VARIABLE:

$$f(y) = f(x)\left|\frac{dx}{dy}\right|$$

TRANSFORMATION OF TWO DIMENSIONAL RANDOM VARIABLE:

$$f(u,v) = f(x,y)\,|J|$$

where $|J| = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$

---

2 Marks:

1. If x has an exponential distribution, with parameter '1' find the p.d.f of $y = \sqrt{x}$.

Solution:

The p.d.f of exponential distribution,

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

Where $\lambda = 1$, $f(x) = e^{-x}$

Given $y = \sqrt{x}$

$y = \sqrt{x}$

$x = y^2$

$\left|\frac{dx}{dy}\right| = 2y$

$\therefore$ p.d.f of y is

$f(y) = f(x)\left|\frac{dx}{dy}\right|$

$f(y) = e^{-x} \cdot 2y$

$f(y) = 2ye^{-y^2}, \; y > 0$

---

2. If x has an exponential distribution with parameter 'λ'. Find p.d.f of $y = \log x$.

Solution:

The p.d.f of exponential distribution,

$f(x) = \lambda e^{-\lambda x}$, $x > 0$

$f(x) = \lambda e^{-\lambda x}, \; x > 0 \; \because \lambda = \lambda$

$y = \log x$

$x = e^y$

$\left|\frac{dx}{dy}\right| = e^y$

P.d.f of y is

$f(y) = f(x)\left|\frac{dx}{dy}\right|$

$= \lambda e^{-\lambda e^y} e^y$

$f(y) = \lambda e^{(y-\lambda e^y)}, y > 0$

$f(y) = \lambda e^{(y-\lambda e^y)}, y>0$

3. If x is uniformly distributed in $(-\pi/2, \pi/2)$. Find p.d.f of $y = \tan x$.

Solution:

The p.d.f of Uniform distribution is,

$f(x) = 1/b-a$

$= \frac{1}{\pi/2 + \pi/2}$

$= 1/\pi$

Given:

$y = \tan x$

$x = \tan^{-1} y$

$\left|\frac{dx}{dy}\right| = \frac{1}{1+y^2}$

The p.d.f of y is

$f(y) = f(x)\left|\frac{dx}{dy}\right|$

$= 1/\pi \cdot 1/1+y^2 = \frac{1}{\pi(1+y^2)}$

$f(y) = \frac{1}{\pi(1+y^2)}, (-\infty < y < \infty)$

4. If x is uniformly distributed in $(-1, 1)$. Find the P.d.f of $y = \sin\frac{\pi x}{2}$

Solution.

The p.d.f of Uniform Distribution is,

$f(x) = 1/_{b-a}$

$= 1/_{1+1}$

$= 1/_2$

Given:

$y = \sin\left(\frac{\pi x}{2}\right)$

$\sin^{-1} y = \frac{\pi x}{2}$

$2/_\pi \sin^{-1} y = x$

$\left|\frac{dx}{dy}\right| = \frac{2}{\pi\sqrt{1-y^2}}$

The p.d.f of y is.

$f(y) = f(x)\left|\frac{dx}{dy}\right|$

$= 1/_2\left(\frac{2}{\pi\sqrt{1-y^2}}\right)$

$f(y) = \frac{1}{\pi\sqrt{1-y^2}} \quad -1<y<1$

1. The Joint p.d.f of 2 dimensional Random variable is given by $f(x,y) = \begin{cases} 4xye^{-(x^2+y^2)} ; & x\geq 0, y\geq 0 \\ 0 ; & \text{otherwise} \end{cases}$. Find the Density Function of $u = \sqrt{x^2+y^2}$.

Solution:

$u = \sqrt{x^2+y^2}, \quad v = y$

$u^2 = x^2+y^2$

$u^2 - y^2 = x^2$

$x = \sqrt{u^2-y^2} ; \quad y = v.$

$x = \sqrt{u^2-v^2} ;$

$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \partial y/_{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$

$= \begin{vmatrix} \frac{2u}{2\sqrt{u^2-v^2}} & \frac{-2v}{2\sqrt{u^2-v^2}} \\ 0 & 1 \end{vmatrix} = \frac{\cancel{2}u}{\cancel{2}\sqrt{u^2-v^2}} = |J|$

$\therefore \quad |J| = \dfrac{u}{\sqrt{u^2-v^2}}$

$\therefore$ The joint p.d.f $f(u,v) = f(x,y)\,|J|$

$f(u,v) = 4xy\, e^{-(x^2+y^2)} \dfrac{u}{\sqrt{u^2-v^2}}$

$= 4\left(\cancel{\sqrt{u^2-v^2}}\right) v e^{-(u^2)} \dfrac{u}{\cancel{\sqrt{u^2-v^2}}}$

$f(u,v) = 4uv\, e^{-u^2}, \quad u \geq 0$

Range space

$x \geq 0; \; y \geq 0$

$\sqrt{u^2-v^2} \geq 0; \; v \geq 0$

$u^2-v^2 \geq 0$

$u^2 \geq v^2$

$u \geq v$

$\therefore \quad u \geq 0, \; 0 \leq v \leq u$

The p.d.f of u is

$f(u) = \int_{-\infty}^{\infty} f(u,v)\,dv$

$= \int_0^u 4uve^{-u^2}\,dv$

$= 4ue^{-u^2}\left(\dfrac{v^2}{2}\right)_0^u$

$= \cancel{4}^{2}ue^{-u^2}\, u^2/\cancel{2}$

$f(u) = 2u^3 e^{-u^2}, \quad u > 0$

---

2. Let x, y be two dimensional Random variable, whose joint p.d.f is given by.

$f(x,y) = e^{-(x+y)}; \; x>0, y>0$

Find p.d.f of $U = \dfrac{x+y}{2}$

Solution:

Let $u = \frac{x+y}{2}$, $x = y$

$2u = x+y$

$x = 2u - y$, $y = v$

$x = 2u - v$

$$|J| = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$= \begin{vmatrix} 2 & -1 \\ 0 & 1 \end{vmatrix}$$

$= |J| = 2$

The p.d.f.

$f(u,v) = f(x,y)\,|J|$

$= e^{-(x+y)}2$

$= 2e^{-(x+y)}$

$f(u,v) = 2e^{-2u}$, $u \geq 0$

Range space:

$x > 0$, $y > 0$

$2u - v > 0$, $v > 0$

$2u > v$

$u > v/2$

$\therefore$ $u \geq 0$, $0 < v < 2u$

The p.d.f of 'u' is

$f(u) = \int_{-\infty}^{\infty} f(u,v)\,dv$

$= \int_{0}^{2u} 2e^{-2u}\,dv$

$= 2e^{-2u}[v]_0^{2u}$

$f(u) = 4ue^{-2u}$

$f(u) = 4ue^{-2u}$, $u \geq 0$

---

3. Let x and y are independent. Given, $f(x) = e^{-x}$, $x>0$, $f(y) = e^{-y}$, $y>0$. Show that $u = x/(x+y)$. $v = x+y$ are independent.

Solution:

Since x & y are independent.

$f(x,y) = f(x) \cdot f(y)$

$f(x,y) = e^{-x}\, e^{-y}$

$f(x,y) = e^{-(x+y)}, \; x>0,\; y>0$

Given:

$U = \dfrac{x}{x+y}$ $\qquad$ $V = x+y$

$y = v - x$

$x = u(x+y)$ $\qquad$ $y = v - uv$

$x = uv$ $\qquad$ $y = v(1-u)$

$$|J| = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$$

$$= \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix}$$

$$= v(1-u) + uv$$

$$= v - \cancel{uv} + \cancel{vu}$$

$|J| = v$

The p.d.f $f(u,v) = f(x,y)\,|J|$

$= e^{-(x+y)}.\,v$

$f(u,v) = v e^{-(v)}$

Range space:

$x>0,\; y>0$

$uv>0,\; v(1-u)>0$

$u>0, v>0.$

$1-u>0.$

$1>u,\; 0<u<1$

The p.d.f of x is.

$f(u) = \int_{-\infty}^{\infty} f(u,v)\,dv$

$= \int_{0}^{\infty} v e^{-v}\,dv$

$$= \left(\frac{ve^{-v}}{-1} - e^{-v} \right)_0^{\infty}$$

$$= 0 - (-1)$$

$f(u) = 1$

The p.d.f of v is.

$$f(v) = \int_{-\infty}^{\infty} f(u,v)\, du$$

$$= \int_0^1 ve^{-v} du = ve^{-v} (u)_0^1$$

$f(v) = ve^{-v}$

$f(u) . f(v) = 1 . ve^{-v} = f(u,v)$

$\therefore$ u and v are independent.

---

4. If the joint p.d.f of 2 dimensional Random. $f(x,y) = x+y$, $0 < x, y < 1$. Find p.d.f of $U = xy$.

Solution:

$u = xy \qquad v = y$

$x = u/y \qquad y = v$

$x = u/v$

$$J = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \partial x/\partial v \\ \frac{\partial y}{\partial u} & \partial y/\partial v \end{vmatrix}$$

$$= \begin{vmatrix} 1/v & -u/v^2 \\ 0 & 1 \end{vmatrix} = 1/v = |J|$$

The p.d.f of U,

$f(u,v) = f(x,y)\ |J|$

$$= (x+y)\left(1/v\right)$$

$$= \left(u/v + v\right)\left(1/v\right)$$

$$= \frac{u+v^2}{v^2} = \frac{u}{v^2} + 1$$

Range space:

$0 < x < 1$; $0 < y < 1$

$0 < u/v < 1$; $0 < u < v$

The p.d.f of U,

$$f(u) = \int_{-\infty}^{\infty} f(u,v)\,dv$$

$$= \int_u^1 \left(\frac{u}{v^2}+1\right) dv = \int_u^1 \left[uv^{-2}+1\right] dv = \left(-uv^{-1}+v\right)_u^1$$

$$= \left(-\frac{u}{v}+v\right)_u^1 = (-u+1) - [-1+u]$$

$$= 1-u+1-u$$

$$= 2-2u$$

$$f(u) = 2(1-u)$$

---

5. If x and y are independent, exponentially distributed with parameter '1'. Find the p.d.f of $U = x-y$.

Solution:

$$f(x,y) = f(x)f(y)$$

$$f(x) = \lambda e^{-\lambda x}$$

$$= e^{-x}$$

$$f(y) = \lambda e^{-\lambda y}$$

$$= e^{-y}$$

$$f(x,y) = e^{-(x+y)}$$

$u = x-y$; $y = v$

$x = y+u$; $v = y$

$x = u+v$

$$|J| = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}$$

$|J| = 1$

$$f(u,v) = f(x,y)\,|J|$$

$$= e^{-(x+y)}(1)$$

$$= e^{-(u+v+v)}(1)$$

$$f(u,v) = e^{-(u+2v)}$$

Range Space:

$x>0$; $y>0$

$u+v>0$; $v>0$

$v > -u$

The p.d.f of U, $\underline{u<0}$

$$f(u) = \int_{-u}^{\infty} e^{-(u+2v)}\,dv$$

$$= e^{-u}\int_{-u}^{\infty} e^{-2v}\,dv$$

$$= e^{-u}\left[\frac{e^{-2v}}{-2}\right]_{-u}^{\infty}$$

$$= e^{-u}\left[0 + \frac{e^{2u}}{2}\right]$$

$$= \frac{e^{u}}{2}$$

$\underline{u>0}$

$$f(u) = \int_{u}^{\infty} e^{-(u+2v)}\,dv$$

$$= e^{-u}\int_{u}^{\infty} e^{-2v}\,dv$$

$$= e^{-u}\left[\frac{e^{-2v}}{-2}\right]_{u}^{\infty}$$

$$= e^{-u}\left[0 + \frac{e^{-2u}}{2}\right]$$

$$= \frac{e^{-u}}{2}$$

CENTRAL LIMIT THEOREM:

If $x_1, x_2, \dots x_n$ be a sequence of independent and identically distributed random variables with $E(x_i) = \mu$ and $var[x_i] = \sigma^2$, $i = 1, 2, 3, \dots n$ if $S_n = x_1 + x_2 + \dots x_n$ then under certain general conditions S_n follows a normal distribution with mean $n\mu$ and variance $n\sigma^2$ as $n \to \infty$.

TYPE - 1:

If the Average of Random variable follows Normal distribution then $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$.

By central limit theorem, $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Problems:

1. The lifetime of a certain brand of an electric bulb may be considered as a random variable with mean 1200 hr and standard deviation 250 hr. Find the probability using central limit theorem that the Average lifetime of 60 bulbs exceeds 1250 hr.

Solution:

Given:

Mean = 1200 hr, μ

Standard deviation, σ = 250hr

No. of samples, $n = 60$

$$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$$

Using central limit theorem,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{\bar{x} - 1200}{250/\sqrt{60}}$$

$z = \dfrac{\bar{x}-1200}{32.275}$

$P(x>1250) = P\left(z > \dfrac{1250-1200}{32.275}\right)$

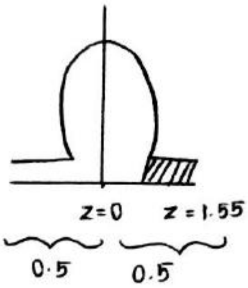
$= P\left(z > {}^{50}/_{32.275}\right)$

$= P(z>1.549)$

$= 0.5 - P(0<z<1.55)$

$= 0.5 - 0.4394$ (From table)

$= 0.0606.$



---

2. A Random sample of size 100 taken from a population whose mean is 60 and variance is 400 using central limit theorem. With what probability can we assert that the mean of sample will not differ from $\mu = 60$ by more than 4.

Solution:

Given:

Mean, $\mu = 60$

variance, $\sigma^2 = 400$

standard deviation, $\sigma = 20$

$n = 100$

$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$

By using central limit theorem,

$z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$z = \dfrac{\bar{x}-60}{20/\sqrt{100}}$

$z = \dfrac{\bar{x}-60}{2}$

$P(|\bar{x}-60| \le 4) = P(-4 \le \bar{x}-60 \le 4)$

$= P(-4+60 \le \bar{x} \le 60+4)$

$= P(+56 \le \bar{x} \le 64)$

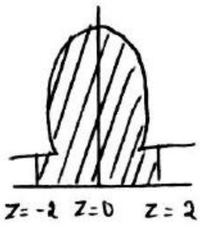
$= P\left(\frac{56-60}{2} \le z \le \frac{64-60}{2}\right)$

$= P(-2 \le z \le 2)$

$= 2P(0 \le z \le 2)$

$= 2(0.4772)$

$= 0.9544$



TYPE-2

If the sum of random variables follows the Normal distribution then S_n follows $N(n\mu, \sigma\sqrt{n})$.

By central limit theorem,

$$z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Problems:

1. If $x_1, x_2, \ldots x_n$ are poisson variance with parameter $\lambda = 2$. use central limit theorem to estimate probability of $(120 \le S_n \le 160)$ Where $S_n = x_1 + x_2 + \ldots x_n$ & $n = 75$.

Solution:

Given:

$n = 75$

In poisson Distribution,

Mean, $\lambda = 2 = \mu$

variance, $\sigma^2 = 2$

standard deviation $\sigma = \sqrt{2}$

$S_n \sim N(n\mu, \sigma\sqrt{n})$

By central limit theorem,

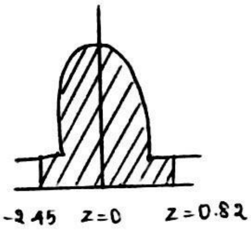
$$z = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$$= \frac{S_n - 150}{\sqrt{150}}$$

To Find $P(120 \le S_n \le 160)$,

$$= P\left(\frac{120-150}{\sqrt{150}} \le z \le \frac{160-150}{\sqrt{150}}\right)$$

$$= P\left(\frac{-30}{\sqrt{150}} \le z \le {}^{10}/_{\sqrt{150}}\right)$$



$$= P(-2.450 \le z \le 0.82)$$

$$= P(-2.45 \le z \le 0) + P(0 \le z \le 0.82)$$

$$= P(0 \le z \le 2.45) + P(0 \le z \le 0.82)$$

$$= 0.4929 + 0.2939 \text{ (From table)}$$

$$= 0.7868.$$

---

2. Let $x_1, x_2 \dots x_{100}$ be independent, identically distributed Random variable (IID) with mean, $\mu = 2$ and $\sigma^2 = 1/4$. Find $(192 < x_1 + x_2 + \dots x_{100} < 210)$

Solution:

Given, $n = 100$

Mean, $\mu = 2$

variance, $\sigma^2 = 1/4$

Standard deviation $(\sigma) = \sqrt{1/4}$

$$= 1/2$$

$S_n \sim N(n\mu, \sigma\sqrt{n})$

By central limit theorem,

$$Z = \frac{S_n - 200}{5}$$

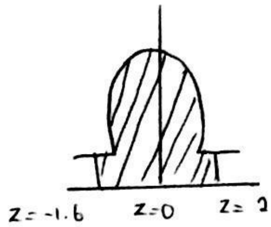
$$P(192 \le S_n \le 210) = P\left(\frac{192-200}{5} \le Z \le \frac{210-200}{5}\right)$$

$$= P(-8/5 \le Z \le 10/5)$$

$$= P(-1.6 \le Z \le 2)$$

$$= P(-1.6 \le Z \le 0) + P(0 \le Z \le 2)$$

$$= P(0 \le Z \le 1.6) + P(0 \le Z \le 2)$$



$$= 0.4452 + 0.4772 \text{ (From table)}$$

$$= 0.9224$$

Type - 3

If the discrete random variable follows normal distribution then $\bar{x}$ follows $N(\mu, \sigma)$.

By central limit theorem,

$$Z = \frac{\bar{x} - \mu}{\sigma}$$

Problem:

1. A coin is tossed 10 times what is the probability of getting 3 or 4 or 5 heads using central limit theorem.

Solution:

In Binomial Distribution,

Mean = np

Here $P = 1/2$, $q = 1/2$

Probability of getting head.

$n = 10$

Mean = $(1/2) 10$

Mean, $\mu = 5$

variance $(\sigma)^2 = npq$
$= (1/2)(1/2)10$

$= (1/4)10$

$= 5/2$

$= 2.5$

standard deviation, $\sigma = \sqrt{2.5} = 1.58$

To approximate the discrete probability distribution to continuous probability distribution, add 0.5 to the upper bound and substract 0.5 from the lower bound.

$P(3-0.5 \le \bar{x} \le 5+0.5) = P(2.5 \le \bar{x} \le 5.5)$

Normal variate, $z = \dfrac{\bar{x}-\mu}{\sigma}$

$z = \dfrac{\bar{x}-5}{1.58}$

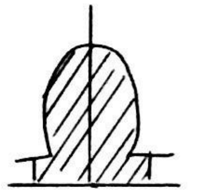
$\therefore P(2.5 \le \bar{x} \le 5.5) = P\left(\dfrac{2.5-5}{1.58} \le z \le \dfrac{5.5-5}{1.58}\right)$

$= P(-1.58 \le z \le 0.32)$

$= P(0 \le z \le -1.58) + P(0 \le z \le 0.32)$

$= 0.4429 + 0.1255$ (From table).

$= 0.5684$



---

1. Three balls are drawn at random without replacement from a box containing 2 white, 3 red and 4 black balls. If x denotes the No. of White balls drawn and y denote the Number of red balls drawn. Find the joint probability distribution of (x, y)

Solution:

Given:

Let x denote No. of White balls drawn

Let y denote No. of red balls drawn

Total = 2+3+4

= 9 balls.

x \ y	0	1	2	3
0	$\frac{(2C_0)(3C_0)(4C_3)}{9C_3} = 1/21$	$\frac{(2C_0)(3C_1)(4C_2)}{9C_3} = 18/24$	$\frac{(2C_0)(3C_2)(4C_1)}{9C_3} = 1/7$	$\frac{(2C_0)(3C_3)(4C_0)}{9C_3} = 1/84$
1.	$\frac{(2C_1)(3C_0)(4C_2)}{9C_3} = 1/4$	$\frac{(2C_1)(3C_1)(4C_1)}{9C_3} = 2/7$	$\frac{(2C_1)(3C_2)(4C_0)}{9C_3} = 1/14$	0 only 3 balls.
2.	$\frac{(2C_2)(3C_0)(4C_1)}{9C_3} = 1/21$	$\frac{(2C_2)(3C_1)(4C_0)}{9C_3} = 1/28$	0	0

Theorem: [Central limit theorem]

If $x_1, x_2, \ldots x_n$ is a sequence of n independent and identically distributed (i.i.d) random variables, each having mean μ and variance σ^2, and if $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$, then the variable $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ has a distribution that approaches the standard normal distribution as $n \to \infty$, provided the m.g.f exits.

Proof:

M.G.F of z about the origin is $M_z(t) = E[e^{tz}]$

$$= E\left[e^{t\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)}\right]$$

$$= E\left[e^{\frac{t\bar{x}\sqrt{n}}{\sigma}} \cdot e^{-\frac{t\mu\sqrt{n}}{\sigma}}\right]$$

$$= e^{-\frac{\mu t\sqrt{n}}{\sigma}} \cdot E\left[e^{\frac{t\sqrt{n}}{\sigma}\left[\frac{x_1+x_2+\dots+x_n}{n}\right]}\right]$$

$$= e^{-\frac{t\mu\sqrt{n}}{\sigma}} \; E\left[e^{\frac{tx_1}{\sigma\sqrt{n}}}, e^{\frac{tx_2}{\sigma\sqrt{n}}} \dots e^{\frac{tx_n}{\sigma\sqrt{n}}}\right]$$

[since $x_1, x_2, \dots x_n$ are independent]

$$E[x_1 x_2 \dots x_n) = [E(x_1) E(x_2) \dots E(x_n)]$$

Hence $M_z(t) = e^{-\frac{t\mu\sqrt{n}}{\sigma}} E\left(e^{\frac{tx_1}{\sigma\sqrt{n}}}\right) E\left(e^{\frac{tx_2}{\sigma\sqrt{n}}}\right) \dots E\left(e^{\frac{tx_n}{\sigma\sqrt{n}}}\right)$

The variables $x_1, x_2 \dots x_n$ have the same M.G.F

$$\therefore M_z(t) = e^{-\frac{\mu t\sqrt{n}}{\sigma}} \left[M_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Where $M_x(t/\sigma\sqrt{n})$ is the m.g.f of $x = x_i, i = 1, 2, 3, \dots n$

Taking log on both sides,

$$\log M_z(t) = \log\left(e^{-\frac{t\mu\sqrt{n}}{\sigma}}\right) + n\log\left(M_x\left(\frac{t}{\sigma\sqrt{n}}\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[E\left(e^{\frac{tx}{\sigma\sqrt{n}}}\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[E\left(1+\left(\frac{t}{\sigma\sqrt{n}}\right)x + \frac{1}{2!}\left(t/\sigma\sqrt{n}\right)^2 x^2 + \dots\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[1+\left(t/\sigma\sqrt{n}\right)\mu_1' + \frac{1}{2!}\left(t/\sigma\sqrt{n}\right)^2\mu_2' + \dots\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\left[\left(\frac{t}{\sigma\sqrt{n}}\mu_1' + \frac{\mu_2'}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \dots\right) - \frac{1}{2}\left(\mu_1'\frac{t}{\sigma\sqrt{n}} + \dots\right)^2 + \dots\right)$$

put $\mu' = \mu = \text{mean}$

$\log M_z(t) = -\frac{\mu t\sqrt{n}}{\sigma} + \frac{\sqrt{n}\mu t}{\sigma} + t^2/2\sigma^2 (\mu_2' - (\mu_1')^2] +$ term containing n in the denominator

$\log M_z(t) = t^2/2\sigma^2 \, \sigma^2 +$ terms containing n ..

$\log M_z(t) = t^2/2$ i.e, $M_z(t) = e^{t^2/2}$ as $n \to \infty$

The M.G.F of z is the m.g.f of $N(0,1)$ ie, as $n \to \infty$ the distribution of z tends to the standard Normal deviation.

UNIT - III

ESTIMATION THEORY

3.1 INTRODUCTION

The problems of statistical inference are divided into problems of estimation and tests of hypotheses. The main difference between these two types is that in problems of estimation we have to determine the value of a parameter or the values of several parameters, from alternatives, whereas in the tests of hypotheses we have to decide whether to accept or reject a specific value or a set of specific values of a parameter. In an estimation problem there is atleast one parameter θ whose value is to be approximated on the basis of a sample. The approximation is performed by using an appropriate statistic. There are two types of estimation procedures.

(i) Point estimation and

(ii) Interval estimation.

3.2 POINT ESTIMATION

Definition: Point Estimator

A statistic used to approximate or estimate a population parameter θ is called a point estimator for θ and is denoted by $\hat{\theta}$.

Definition: Point Estimate

The numerical value assumed by the statistic when evaluated for a given sample is called a point estimate for θ.

Example:

If we use a value of $\overline{X}$ to estimate the mean of a population, an observed sample proportion to estimate the parameter θ of a binomial population or a value of S^2 to estimate a population variance using a point estimate of the parameter.

These estimates are called point estimates because in each case a single number or a single point on the real axis, is used to estimate the parameter.

Note that there is a difference between the terms estimator and estimate. The estimator is the statistic used to generate the estimate and it is a random variable whereas an estimate is a number.

Since estimators are random variables, the problem of point estimation is to study their sampling distributions. For example, when we estimate the variance of a population on the basis of a random sample, we expect that the values of S^2 equal to σ^2, but to know whether we can expect it to be close. Also we have to decide, whether to use a sample mean or a sample median to estimate the mean of a population, whether $\overline{X}$ or $\tilde{X}$ is more likely to yield a value that is actually close.

Various statistical properties of estimators used to decide which estimator is most appropriate in a given situation are unbiasedness, minimum variance, efficiency, consistency, sufficiency and robustness.

Definition: Unbiased estimator

A statistic or point estimator $\hat{\theta}$ is said to be an unbiased estimator or its value be an unbiased estimate, if and only if the mean of the sampling distribution of the estimator is equal to θ.

$$\text{(ie)} \quad E[\hat{\theta}] = \theta.$$

Definition: Biased estimator

If the estimator is not unbiased, then $E[\hat{\theta}] - \theta$ is called the biased estimator of the estimate θ. That means if the estimator is unbiased then $E[\hat{\theta}] - \theta = 0$.

Hence, a statistic is unbiased, if the expected value (Average value) should be equal to the parameter which is supposed to estimate.

3.3 MORE EFFICIENT UNBIASED ESTIMATOR

Definition :

A statistic $\hat{\theta}_1$ is said to be a more efficient unbiased estimate of the parameter θ than the statistic $\hat{\theta}_2$ if

(i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimates of θ.

(ii) The variance of the sampling distribution of the first estimator $\hat{\theta}_1$ is less than that of the second estimator $\hat{\theta}_2$.

3.4 MAXIMUM ERROR OF ESTIMATE

We know that for random samples from normal population, the mean is more efficient than the median as an estimate of μ, when we estimate a population mean μ, the variance of sampling distribution of no other statistic is less than that of the sampling distribution of the mean. When we use a sample mean to estimate the mean of a population, together with method of estimation which has some properties, that the estimate equals μ. Hence to accompany such a point estimate of μ with statements as how close we expect the estimate to be. Then the error $\bar{x}-\mu$ is the difference between the estimate and the quantity to estimate. To examine this error, for large n, $\dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is a value of a random variable having the standard normal distribution.

$$P\left[-Z_{\alpha/2} \le \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \le Z_{\alpha/2}\right] = 1-\alpha.$$

$$P\left[\frac{|\bar{x}-\mu|}{\left(\frac{\sigma}{\sqrt{n}}\right)} \le Z_{\alpha/2}\right] = 1-\alpha$$

(or) $$P\left[|\bar{x}-\mu| \le Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

Here $Z_{\alpha/2}$ is the normal curve area to its right equals $\alpha/2$. It is noted that $|\bar{x}-\mu|$ is the error in estimating μ by the unbiased estimator of the sample mean $\bar{x}$. Let E denote the maximum value of $|\bar{x}-\mu|$, then

$E = Z_{\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}$ (Large samples, σ known) with probability $1-\alpha$. That mean, if we want to estimate μ with the mean of a large sample $(n \geq 30)$ we can assert with probability $1-\alpha$ that the error $|\bar{x}-\mu|$ will be atmost $Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. The common values for $1-\alpha$ are 0.95 (5% level) and 0.99 (1% level) and the corresponding values of $Z_{\alpha/2}$ are $Z_{0.025} = 1.96$ and $Z_{0.005} = 2.575$ respectively.

The formula for finding the value of E can also be applied to determine the sample size to get the desired degree of accuracy. Suppose we use the mean of a large random sample to estimate the mean of a population and to assert with the probability $(1-\alpha)$ that the error would be the quantity E. Then the sample size can be computed by using the formula

$$n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E}\right]^2.$$

To apply this formula, we must know the values of $1-\alpha$, E and σ.

For small samples when σ is unknown then let us consider $|t| = \frac{\bar{X}-\mu}{\left(\frac{s}{\sqrt{n}}\right)}$ with $(n-1)$ degrees of freedom.

Hence the maximum error of estimate for small sample when σ is unknown, is given by

$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ (small samples, σ unknown)

WORKED EXAMPLES

Example: 1

If X has the binomial distribution with parameters n and θ, show that the sample proportion, $\frac{X}{n}$ is an unbiased estimator of θ.

Solution:

We know that the probability mass function of Binomial distribution is

$$P(x=X) = nC_x\, p^x\, q^{n-x};\ x = 0, 1, 2 \ldots$$

The mean of the binomial distribution is $E[X] = np$ where n and p are parameters.

Since $E[X] = n\theta$ (p is replaced by θ),

$$E\left[\frac{X}{n}\right] = \frac{1}{n} E[X] = \frac{1}{n} \cdot n\theta = \theta.$$

Hence $\frac{X}{n}$ is an unbiased estimator of θ.

Example: 2

If $X_1, X_2, X_3 \ldots X_n$ constitute a random sample from the population given by

$$f(x) = \begin{cases} e^{-(x-\delta)} & \text{for } x > \delta \\ 0 & \text{otherwise} \end{cases}$$

Show that $\overline{X}$ is a biased estimator of δ.

Solution:

The mean of the population is given by

$$\overline{X} = E[X] = \mu = \int_{\delta}^{\infty} x \cdot e^{-(x-\delta)}\, dx.$$

By using Bernoulli's formula for integration we get

$\int uv\,dx = uv_1 - u' v_2 + u'' v_3 - u''' v_4 + \ldots$

$$E[X] = \left[x\left\{\frac{e^{-(x-\delta)}}{-1}\right\} - (1)\left\{\frac{e^{-(x-\delta)}}{1}\right\}\right]_{\delta}^{\infty}$$

$$= \left[-xe^{-(x-\delta)} - e^{-(x-\delta)}\right]_{\delta}^{\infty}$$

$$= [\{0 - 0\} - \{-\delta e^{-0} - e^{-0}\}]$$

$$= [1 + \delta]$$

It follows that $E[X] = \overline{X} = 1 + \delta \neq \delta$.

Hence $\overline{X}$ is a biased estimator of δ.

Example: 3

A random variable has the binomial distribution and get x success in n trials, show that $\dfrac{x+1}{n+2}$ is not an unbiased estimate of the binomial parameter p.

✍ **Solution:**

We know that the mean of the binomial distribution is $E[X] = \overline{X} = np$ where n and p are parameters.

$$E[X+1] = E[X] + E[1] = np + 1$$

$$\therefore\ E\left[\frac{X+1}{n+2}\right] = \frac{1}{n+2} E[X+1] = \frac{1}{n+2}(np+1)$$

$$\therefore\quad \frac{np+1}{n+2} \neq p$$

Hence $\dfrac{X+1}{n+2}$ is not an unbiased estimate of p.

Example: 4

Let $y_1, y_2, y_3 \ldots y_n$ be random variables with mean m. The quantity $\bar{y} = \dfrac{1}{n}\sum\limits_{i=1}^{n} y_i$ is the sample mean. Verify that whether it is unbiased or not.

✍ **Solution:**

Let us consider

$$E[\bar{y}] = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[y_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} m = \frac{1}{n}\cdot nm = \frac{nm}{n} = m.$$

$\therefore\ \bar{y}$ is an unbiased estimator of m.

Example: 5

Let $y_1, y_2, y_3 \cdots y_n$ be scalar random variables independent and identically distributed with mean m and variance σ^2. Verify the given quantity $\hat{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$ is unbiased or not for the variance σ^2.

Solution:

We know that from the above Example 4.

$$\bar{y} = \frac{1}{n} \sum_{j=1}^{n} y_j$$

$$\therefore \; E[\hat{\sigma}_y^2] = E\left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\left[\left(y_i - \frac{1}{n} \sum_{j=1}^{n} y_j\right)^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n^2} E\left[\left(n y_i - \sum_{j=1}^{n} y_j\right)^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{n^2} E\left[\left\{n(y_i - m) - \sum_{j=1}^{n} (y_j - m)\right\}^2\right]$$

Now let us consider

$$E\left[\left\{n(y_i - m) - \sum_{j=1}^{n} (y_j - m)\right\}^2\right]$$

$$= E\left[n^2 (y_i - m)^2 + \sum_{j=1}^{n} (y_j - m)^2 - 2n(y_i - m) \sum_{j=1}^{n} (y_j - m)\right]$$

$$= n^2 E[(y_i - m)^2] + E\left[\left(\sum_{j=1}^{n} (y_j - m)\right)^2\right] - 2n\, E\left[(y_i - m)\sum_{j=1}^{n} (y_j - m)\right]$$

$$= n^2 \sigma^2 + n\,\sigma^2 - 2n\,\sigma^2$$

$$= n^2 \sigma^2 - n\,\sigma^2$$

$$= n\,(n-1)\,\sigma^2$$

$$\therefore\; E[\hat{\sigma}_y^2] = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{n^2}\, n\,(n-1)\,\sigma^2$$

$$= \frac{1}{n^3}\, n\,(n-1)\left[\sum_{i=1}^{n} \sigma^2\right]$$

$$= \frac{1}{n^3}\, n\,(n-1)\,(n\,\sigma^2)$$

$$= \frac{(n-1)}{n}\,\sigma^2$$

$$\neq \sigma^2$$

Hence $\hat{\sigma}_y^2$ is not an unbiased estimate for the variance σ^2.

Example: 6

Let $y_1, y_2, y_3 \ldots y_n$ be independent and identically distributed scalar random variables, with mean m and variance σ^2. The quantity $S^2 = \dfrac{1}{(n-1)}\sum_{i=1}^{n} (y_i - \bar{y})^2$ is called sample variance. Verify for unbiasedness.

Solution:

It is given that

$$S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{n}{n-1}\,\hat{\sigma}_y^2$$

$$E[S^2] = \frac{n}{n-1} E[\hat{\sigma}_y^2] = \frac{n}{n-1} \cdot \frac{1}{n} (n-1)\sigma^2 = \sigma^2.$$

$\therefore$ S^2 is an unbiased estimator of the variance σ^2.

Example: 7

Let $x_1, x_2, x_3 \ldots x_n$ be a random sample from a normal population $N(\mu, 1)$. Show that $t = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ is an unbiased estimator of $1 + \mu^2$.

Solution:

It is given that $N(\mu, 1)$.

That means the mean is μ and variance is 1 in the standard normal population.

$\therefore$ $E(X_i) = \mu$ and $\text{Var}[X_i] = 1 \;\forall\; i = 1, 2 \ldots n$

We know that $\text{Var}[X_i] = E[X_i^2] - \{E[X_i]\}^2$

Now

$$E[X_i^2] = \text{Var}[X_i] + [E(X_i)]^2$$

$$E[X_i^2] = 1 + \mu^2$$

$$E[t] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i^2]$$

$$= \frac{1}{n}\sum_{i=1}^{n} [1 + \mu^2] = \frac{1}{n} \cdot n[1 + \mu^2]$$

$$= 1 + \mu^2$$

Hence t is an unbiased estimator of $1 + \mu^2$.

Example: 8

Let $x_1, x_2, x_3 \ldots x_n$ be random samples on a Bernoulli variable X taking the value 1 with probability θ and the value with 0 with probability $(1-\theta)$. Show that $\frac{\tau(\tau-1)}{n(n-1)}$ is an unbiased estimate of θ^2 where $\tau = \sum_{i=1}^{n} X_i$.

Solution:

Since X_i takes only the values 1 and 0 with respective probabilities θ and $(1-\theta)$ we have

$$E[X_i] = 1 \cdot \theta + 0(1-\theta) = \theta$$

$$E[X_i^2] = 1^2 \cdot \theta + 0^2(1-\theta) = \theta$$

$$\text{Var}[X_i] = E[X_i^2] - \{E[X_i]\}^2$$

$$= \theta - \theta^2$$

$$= \theta(1-\theta)$$

$$E(\tau) = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \theta = n\theta.$$

$$\text{Var}(\tau) = \text{Var}[X_1 + X_2 + X_3 + \ldots X_n]$$

$$= \text{Var}[X_1] + \text{Var}[X_2] + \ldots + \text{Var}[X_n]$$

The covariance terms vanish since $x_1, x_2, x_3 \ldots x_n$ are independent.

$$\text{Var}[\tau] = \text{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \text{Var}[X_i]$$

$$= \sum_{i=1}^{n} \theta(1-\theta) = n\theta(1-\theta).$$

$$E[\tau^2] = \text{Var}[\tau] + \{E[\tau]\}^2$$
$$= n\theta(1-\theta) + n^2\theta^2$$
$$E[\tau^2] = n\theta[1-\theta+n\theta]$$

Now $$E\left[\frac{\tau(\tau-1)}{n(n-1)}\right] = \frac{1}{n(n-1)} E[\tau(\tau-1)]$$
$$= \frac{1}{n(n-1)}[E(\tau^2) - E(\tau)]$$
$$= \frac{1}{n(n-1)}[n\theta(1-\theta+n\theta) - n\theta]$$
$$= \frac{1}{n(n-1)}[n\theta - n\theta^2 + n^2\theta^2 - n\theta]$$
$$= \frac{1}{n(n-1)}[n\theta^2(n-1)]$$
$$= \theta^2$$

Hence $\frac{\tau(\tau-1)}{n(n-1)}$ is an unbiased estimate of θ^2.

Example: 9

In a company, an engineer wishes to apply the mean of a random sample of size $n = 150$ (large sample) to estimate the average mechanical aptitude of assembly line workers. Based on his experience, the engineer assumes that $\sigma = 6.2$ for such date. What does he assert with probability 0.99 about the maximum size of his error?

Solution:

It is given that $n = 150, \sigma = 6.2$ and $Z_{0.005} = 2.575$.

We know that the maximum error of estimate for large sample when σ known is given by

$$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 2.575\left(\frac{6.2}{\sqrt{150}}\right) = 1.30.$$

Hence the engineer asserts with probability 0.99 that this maximum error of estimate is 1.30.

Example: 10

A machine worker wishes to determine the average time it takes a mechanic to rotate the tires of a lorry, and he wants to assert with 95% confidence that the mean of his sample is off by atmost 0.50 minute. If he assumes from his past experience that $\sigma = 1.6$ minutes, how large a sample will he has to take?

Solution:

It is given that $E = 0.50$, $\sigma = 1.6$ and $Z_{0.025} = 1.96$. Since the size of the sample is not known, we have to find the size of the sample first.

$$\therefore \quad n = \left[\frac{Z_{\alpha/2} \cdot \sigma}{E}\right]^2 = \left[\frac{1.96 \times 1.6}{0.50}\right]^2 = 39.337984$$

$\therefore \; n = 40$

Hence, the worker will have to take 40 mechanics to perform the task of rotating the tires of a lorry.

Example: 11

In 6 determinations of the melting point of bowl, a chemist obtained a mean of 232.26°C with a S.D of 0.14°C. If he uses this mean as the actual melting point of bowl, what can the chemist assert with 98% confidence about the maximum error?

Solution:

It is given that $n = 6, s = 0.14, t_{0.01} = 3.365$ for $n - 1 = 5$ degrees of freedom. Then we know that

$$E = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 3.365\left(\frac{0.14}{\sqrt{6}}\right) = 0.19\,.$$

Hence, the chemist asserts that with 98% confidence that his value of the melting point of bowl is off by atmost 0.19 degree.

3.5 INTERVAL ESTIMATION

Using point estimation, sometimes we may not get desired degree of accuracy in estimating parameter. Hence by replacing the point estimation by interval estimation, we can assert with reasonable degree of certainity that they will contain the parameter under consideration.

Definition:

The interval estimate of an unknown parameter θ is an interval of the form $L \le \theta \le U$. Here the end points L and U depend on the numerical value of the statistic $\hat{\theta}$ for a sample on the sampling distribution of $\hat{\theta}$.

Note:

The advantage of an interval estimate over a point estimate is that the interval estimate is formulated in such a way that we can assess the confidence that the interval contains the parameter. The interval estimators are called confidence intervals.

Definition: Confidence Interval

The $100\,(1-\alpha)\,\%$ confidence interval for the parameter θ is in the form of $[L, U]$ such that $P\,[L \le \theta \le U] = 1-\alpha,\ 0<\alpha<1$. Here L and U are called the lower and upper confidence limits respectively $(1-\alpha)$ is the confidence coefficient or the degree of confidence. When $\alpha = 0.01$, the confidence coefficient is 0.99 and it has 99% confidence interval.

3.5.1 Confidence interval for the mean when σ is known

Suppose that we have a large $(n \ge 30)$ random sample from a population with unknown mean μ and known variance σ^2.

For large n, $Z = \dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$ a random variable is having the standard normal distribution.

$$\therefore P\,[-Z_{\alpha/2} \le Z \le Z_{\alpha/2}] = 1-\alpha$$

$$P\left[-Z_{\alpha/2} \le \frac{\overline{X}-\mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} \le Z_{\alpha/2}\right] = 1-\alpha$$

$$\therefore\ P\left[\overline{X} - Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

3.5.2 Large sample confidence interval for μ, σ known

Definition:

If $\bar{x}$ is the sample mean of a random sample of size n from a population with known σ^2, the $100(1-\alpha)$ % confidence interval on μ is given by

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Note:

The above confidence interval formula is applicable only for random samples from normal populations for large samples.

3.5.3 Small sample confidence interval for μ, σ unknown

Definition:

For small samples $(n < 30)$ and the sample is from normal population, we have to use t-distribution.

If $\bar{x}$ and s are the mean and S.D of a random sample from a normal distribution respectively with unknown variance σ^2, then the confidence interval is

$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ with $n-1$ degrees of freedom in t distribution.

WORKED EXAMPLES

Example: 1

A random sample of size $n = 100$ is taken from a population with $\sigma = 5.1$, $\bar{x} = 2.16$. Construct a 95% confidence interval for the population mean μ.

Solution:

Give that $n = 100$, $\sigma = 5.1$, $\bar{x} = 21.6$,

$$1 - \alpha = 0.95, \alpha = 0.05 \text{ and } Z_{\alpha/2} = Z_{0.025} = 1.96.$$

We know that for large sample ($n = 100$) the confidence interval for μ when σ known is

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$21.6 - 1.96\left(\frac{5.1}{\sqrt{100}}\right) < \mu < 21.6 + 1.96\left(\frac{5.1}{\sqrt{100}}\right)$$

$$\Rightarrow 20.6 < \mu < 22.6$$

Thus we can assert with 95% confidence that the mean (μ) lies in the interval (20.6, 22.6).

Example: 2

Construct a 99% confidence interval for the mean given that $n = 80$, $\bar{x} = 18.85$ and $s^2 = 30.77$.

Solution:

Given that $n = 80, \bar{x} = 18.85$, $s^2 = 30.77$ then $s = 5.55$.

We know that

$$\bar{x} - Z_{\alpha/2} \cdot \left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{x} + Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$\Rightarrow 18.85 - 2.575\left(\frac{5.55}{\sqrt{80}}\right) < \mu < 18.85 + 2.575\left(\frac{5.55}{\sqrt{80}}\right)$$

$$\Rightarrow 17.25 < \mu < 20.45$$

It is 99% confident that the interval from 17.25 to 20.45 contains the average μ.

Example: 3

The mean weight loss of $n = 16$ grinding balls after a certain length of time in mill slurry is 3.42 grams with a S.D of 0.68 grams. Construct 99% confidence interval for the true mean weight loss of such grinding balls under the given conditions.

Solution:

Since $n = 16$, it is belonging to small sample.

Also it is given that $n = 16, \bar{x} = 3.42, s = 0.68$ and $t_{0.005} = 2.947$ for $n-1 = 15$ degrees of freedom for μ, we have

$$\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

(ie) $3.42 - 2.947\left(\frac{0.68}{\sqrt{16}}\right) < \mu < 3.42 + 2.947\left(\frac{0.68}{\sqrt{16}}\right)$

$\Rightarrow 2.92 < \mu < 3.92.$

We have 99% confident that the interval from 2.92 to 3.92 contains the mean weight loss.

Theorem 1: If S^2 is the variance of a random sample from an infinite population with finite variance σ^2, then $E[S^2] = \sigma^2$.

Proof: We know that, if $X_1, X_2, X_3 \dots X_n$ constitute a random sample, then $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is called the sample mean and

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \text{ is called the sample variance.}$$

Then $E[S^2] = E\left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2\right]$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (X_i - \bar{X} + \mu - \mu)^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} \{(X_i - \mu) - (\bar{X} - \mu)\}^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} \{(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i - \mu)^2 - 2E\sum_{i=1}^{n} (X_i - \mu)(\bar{X} - \mu) + E\sum_{i=1}^{n} (\bar{X} - \mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i-\mu)^2 - \sum_{i=1}^{n} E(\bar{X}-\mu)^2\right]$$

We know that $E[(X_i-\mu)^2] = \sigma^2$ and $E[(\bar{X}-\mu)^2] = \frac{1}{n}\sigma^2$

$$\therefore E[S^2] = \frac{1}{n-1}\left[\sum_{i=1}^{n} \sigma^2 - \sum_{i=1}^{n} \frac{1}{n}\sigma^2\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - n\cdot\frac{1}{n}\sigma^2\right]$$

$$= \frac{1}{n-1}[n\sigma^2 - \sigma^2]$$

$$= \frac{1}{n-1}(n-1)\sigma^2$$

$$E[S^2] = \sigma^2$$

3.6 EFFICIENCY

If we select one of the several unbiased estimators of a given parameter, we select the one whose sampling distribution has the smallest variance. To verify whether a given unbiased estimator has the smallest variance, whether it is a minimum variance unbiased estimator (also called a best unbiased estimator), we can use the fact that if $\hat{\theta}$ is an unbiased estimator of θ, that the variance of $\hat{\theta}$ must satisfy the inequality

$$\text{Var}[\hat{\theta}] \geq \frac{1}{n\cdot E\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]}$$

where $f(x)$ is the value of the population density at x and n is the size of the random sample. This inequality is called the Cramer - Rao inequality.

Theorem 2: If $\hat{\theta}$ is an unbiased estimator of θ, and

$$\text{Var}[\hat{\theta}] = \frac{1}{n\cdot E\left[\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2\right]}$$

then $\hat{\theta}$ is a minimum variance unbiased estimator of θ.

Example: 4

A sample of size 25 from a normal population with variance 81, produced a mean of 81.2. Find a 0.95 level of confidence interval for the mean.

Solution:

We know that the confidence interval for the mean when σ is known is given by

$$P\left[\overline{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha.$$

Then it is given that $\overline{X} = 81.2, \sigma^2 = 81, n = 25$ and $Z_{\alpha/2} = 1.96$. Since $\sigma^2 = 81$; $\sigma = 9$.

$$81.2 - 1.96 \frac{9}{\sqrt{25}} < \mu < 81.2 + 1.96 \frac{9}{\sqrt{25}}$$

$$81.2 - 1.96 \frac{9}{5} < \mu < 81.2 + 1.96 \frac{9}{5}$$

$\Rightarrow 81.2 - 3.525 < \mu < 81.2 + 3.525$

$\Rightarrow 77.675 < \mu < 81.725$

$\Rightarrow (77.675, 81.725)$

Example: 5

Show that $\overline{X}$ is a minimum variance unbiased estimator of the mean μ of a normal population.

Solution:

We know that the probability density function of normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < \infty.$$

Take log on both sides

$$\ln f(x) = \ln\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right]$$

$$\ln f(x) = \ln\left[\frac{1}{\sigma\sqrt{2\pi}}\right] + \ln\left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right]$$

$$\ln f(x) = -\ln\sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \qquad \text{... (1)}$$

$$\left(\because \ln\left(\frac{A}{B}\right) = \ln A - \ln B,\ \ \ln 1 = 0, \text{ and } \ln e^x = x\right)$$

Differentiate equation (1) partially w.r.to μ

$$\frac{\partial}{\partial\mu}[\ln f(x)] = \frac{\partial}{\partial\mu}\left[-\ln\sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$= -\frac{1}{2}\frac{\partial}{\partial\mu}\left[\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$= -\frac{1}{2}\cdot\frac{1}{\sigma^2}\cdot 2(x-\mu)(-1)$$

$$= \frac{x-\mu}{\sigma^2} = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)$$

Then

$$E\left[\left\{\frac{\partial}{\partial\mu}(\ln f(x))\right\}^2\right] = E\left[\frac{1}{\sigma^2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$= \frac{1}{\sigma^2}E\left[\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$= \frac{1}{\sigma^2}\cdot 1 = \frac{1}{\sigma^2}$$

$$\left(\because E\left[\left(\frac{x-\mu}{\sigma}\right)^2\right] = 1\right)$$

$$\therefore \frac{1}{n \cdot E\left[\left\{\frac{\partial}{\partial \mu}[\ln f(x)]\right\}^2\right]} = \frac{1}{n \cdot \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

and since $\overline{X}$ is unbiased and Var $[\overline{X}] = \frac{\sigma^2}{n}$, it follows that $\overline{X}$ is a minimum variance unbiased estimator of μ.

Definition: Most efficient estimator

If in a class of consistent estimators for a parameter, there exists one whose sampling variance is less than that of any such estimator, it is called the most efficient estimator. Whenever such an estimator exists, it provides a criterion for measurement of efficiency of the other estimators.

Definition: Efficiency

If $\hat{\theta}_1$ is the most efficient estimator with variance v_1 and $\hat{\theta}_2$ is any other estimator with variance v_2, then the efficiency E of $\hat{\theta}_2$ is defined as

$$E = \frac{v_1}{v_2}$$

Here E cannot exceed unity.

3.7 CONSISTENCY

In the preceeding section, we assumed that the variance of an estimator or its mean square error, is a good sign of its chance fluctuations. The fact that these measures may not provide good criteria for this purpose. For large n, the estimators will take on values that are very close to the respective parameters.

Definition: Consistency

The statistic $\hat{\theta}$ is a consistent estimator of the parameter θ if and only if for each $c > 0$

$$\lim_{n \to \infty} P[|\hat{\theta} - \theta| < c] = 1.$$

Consistency is an asymptotic property. That means limiting property of an estimator. When n is large, the error made with a consistent estimator will be less than any small preassigned positive constant. The kind of convergence expressed by the limit in the above definition is called convergence in probability.

Theorem 3: If $\hat{\theta}$ is an unbiased estimator of the parameter θ and $\text{Var}[\hat{\theta}] \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ.

Example: 6

Show that for a random sample from a normal population, the sample variance S^2 is a consistent estimator of σ^2.

Solution:

We know that if S^2 is the variance of a random sample from an infinite population with the finite variance σ^2, then $E(S^2) = \sigma^2$. Since S^2 is an unbiased estimator of σ^2 it is obvious that $\text{Var}[S^2] \to 0$ as $n \to \infty$.

Also we know that if $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with mean μ and S.D σ, then

(i) $\overline{X}$ and S^2 are independent.

(ii) The random variable $\frac{(n-1)}{\sigma^2} S^2$ has a Chi-square distribution with $n-1$ degrees of freedom.

From the above definition, we find that for a random sample from a normal population.

$$\text{Var}[S^2] = \frac{2\sigma^4}{n-1}.$$

It follows that $\text{Var}[S^2] \to 0$ as $n \to \infty$, and thus S^2 is a consistent estimator of the variance of a normal population.

Example: 7

Prove that in sampling from a normal population $N(\mu, \sigma^2)$, the sample mean is consistent estimator of μ.

Solution:

In sampling from a $N(\mu, \sigma^2)$ population, the sample mean $\bar{x}$ is also normally distributed as $N\left(\mu, \frac{\sigma^2}{n}\right)$.

$$\therefore\ E[\bar{X}] = \mu \text{ and } \operatorname{Var}[\bar{X}] = \frac{\sigma^2}{n}.$$

Hence as $n \to \infty$; $E(\bar{X}) = \mu$ and $\operatorname{Var}[\bar{X}] = 0$.

Hence by Theorem 3, $\bar{X}$ is a consistent estimator of μ.

3.8 SUFFICIENCY

An estimator $\hat{\theta}$ is said to be sufficient if it contains all the information in a sample relevant to the estimation of θ. (ie) If all the knowledge about θ that can be gained from an individual sample values and their order can be gained from the value of $\hat{\theta}$ alone.

Definition:

The statistic $\hat{\theta}$ is a sufficient estimator of the parameter θ if and only if for each value of $\hat{\theta}$, the conditional probability distribution or density of the random sample $X_1, X_2 \ldots X_n$ given $\hat{\theta}$ is independent of θ.

Definition:

A random variable X has a Bernoulli distribution and it is referred to as a Bernoulli random variable, it and only if its probability distribution is given by

$$f(x, \theta) = \theta^x (1-\theta)^{1-x} \text{ for } x = 0, 1.$$

Example: 8

If $X_1, X_2, X_3 \ldots X_n$ constitute a random sample of size n from a Bernoulli population show that

$$\hat{\theta} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}$$

is a sufficient estimator of the parameter θ.

Solution:

By the definition of Bernoulli distribution we know that

$$f(x\,;\;\theta) = \theta^x (1-\theta)^{1-x} \text{ for } x = 0, 1.$$

Now $f(x_i\,;\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$; $x = 0, 1$ and $i = 1, 2, 3 \ldots n$.

$$\Rightarrow f(x_1, x_2, x_3 \ldots x_n) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$$

$$= \theta^x (1-\theta)^{n-x}$$

$$= \theta^{n\hat{\theta}} (1-\theta)^{n-n\hat{\theta}} \text{ for } x_i = 0 \text{ or } 1 \text{ and } i = 1, 2, 3 \ldots n.$$

Also since $X = X_1 + X_2 + X_3 \ldots + X_n$ is a binomial random variable with parameters θ and n, its distribution is given by

$$b\,(x\,;\;n, \theta) = nC_x\,\theta^x (1-\theta)^{n-x}$$

and the transformation of variable technique, we have

$$g\,(\hat{\theta}) = nC_{n\hat{\theta}}\;\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \text{ for } \hat{\theta} = 0, \frac{1}{n}, \ldots 1$$

Then $\dfrac{f(x_1, x_2, x_3 \ldots, x_n\,;\;\hat{\theta})}{g\,(\hat{\theta})} = \dfrac{f(x_1, x_2, x_3 \ldots x_n)}{g\,(\hat{\theta})}$

$$= \frac{\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}}{nC_{n\hat{\theta}}\,\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}} = \frac{1}{nC_{n\hat{\theta}}}$$

$$= \frac{1}{nC_x}$$

$$= \frac{1}{nC_{x_1 + x_2 + x_3 + \dots x_n}} \text{ for } x_i = 0 \text{ or } 1$$

and $i = 1, 2, 3, 4 \dots n$.

This does not depend on θ and that $\hat{\theta} = \frac{X}{n}$ is a sufficient estimator of θ.

Example: 9

Show that $Y = \frac{1}{6}[X_1 + 2X_2 + 3X_3]$ is not a sufficient estimator of Bernoulli parameter θ.

Solution:

Since we must show that

$$f(x_1, x_2, x_3/y) = \frac{f(x_1, x_2, x_3, y)}{g(y)}$$

is not independent of θ for some values of X_1, X_2 and X_3.

Let us consider $X_1 = 1, X_2 = 1$ and $X_3 = 0$.

$$\therefore\ Y = \frac{1}{6}[X_1 + 2X_2 + X_3] = \frac{1}{6}[1 + 2.1 + 3.0]$$

$$= \frac{3}{6} = \frac{1}{2}$$

$$\text{Now } f\left[1, 1, 0/Y = \frac{1}{2}\right] = \frac{P\left[X_1 = 1, X_2 = 1, X_3 = 0, Y = \frac{1}{2}\right]}{P\left[Y = \frac{1}{2}\right]}$$

$$= \frac{f(1, 1, 0)}{f(1, 1, 0) + f(0, 0, 1)}$$

We know that $f(x\,; \theta) = \theta^x (1 - \theta)^{1-x}$ for $x = 0, 1$.

$f(x_1, x_2, x_3) = \theta^{x_1 + x_2 + x_3} (1 - \theta)^{3 - (x_1 + x_2 + x_3)}$ for $x_i = 0, 1$ and $i = 1, 2, 3$.

Here

$$f(1, 1, 0) = \theta^{1+1+0} (1 - \theta)^{3 - (1+1+0)} = \theta^2 (1 - \theta)$$

$$f(0, 0, 1) = \theta^{0+0+1} (1 - \theta)^{3 - (0+0+1)} = \theta(1 - \theta)^2$$

$$\therefore f\left(1, 1, 0/Y = \frac{1}{2}\right) = \frac{f(1, 1, 0)}{f(1, 1, 0) + f(0, 0, 1)}$$

$$= \frac{\theta^2 (1-\theta)}{\theta^2 (1-\theta) + \theta (1-\theta)^2}$$

$$= \frac{\theta^2 (1-\theta)}{(1-\theta)\,[\theta^2 + \theta (1-\theta)]}$$

$$= \frac{\theta^2}{(\theta^2 + \theta - \theta^2)} = \theta .$$

Hence it can be seen that this conditional probability depends on θ. Thus it is shown that $Y = \frac{1}{6}[X_1 + 2X_2 + 3X_3]$ is not a sufficient estimator of the parameter θ of a Bernoulli population.

Theorem 4: The statistic $\hat{\theta}$ is a sufficient estimator of the parameter θ if and only if the joint probability distribution or density of the random sample can be factored so that

$$f(x_1, x_2, x_3 \dots x_n ;\ \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, x_3 \dots x_n),$$

where $g(\hat{\theta}, \theta)$ depends only on $\hat{\theta}$ and θ and $h(x_1, x_2, x_3 \dots x_n)$ does not depend on θ.

Example: 10

Show that $\overline{X}$ is a sufficient estimator of the mean μ of a normal population with known variance σ^2.

Solution:

We know that the probability density function of a normal distribution is

$$f(x ;\ \mu) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

By making use of the above fact, we have

$$f(x_1, x_2, x_3 \dots x_n ; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Let $\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} [x_i - \mu + \bar{x} - \bar{x}]^2$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (\bar{x} - \mu)^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

We get

$$f(x_1, x_2, x_3 \dots x_n ; \mu) = \left\{\frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right)^2}\right\} \times$$

$$\left\{\frac{1}{\sqrt{n}}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}\right\}.$$

Here the first term on the R.H.S depends only on the estimate $\bar{x}$ and the population mean μ, whereas the second term on R.H.S does not depend μ. Hence by the above theorem 3, it follows that $\bar{X}$ is a sufficient estimator of the mean μ of a normal population with known variance σ^2.

3.9 ROBUSTNESS

One of the important statistical properties is robustness. Robustness is an indicative of the extent to which estimation procedures are adversely affected by violations of underlying assumptions. That means, an estimator is said to be robust if its sampling distribution is not affected by violations of assumptions. Such violations are due to out liers caused by outright errors made by reading instruments or recording the data or by mistakes in experimental procedures. They may depend on the nature of the populations sampled or their parameters.

For example, when estimating the average life of an electric component, we think that, we are sampling an exponential population, whereas actually we are sampling a Weibull population, or when estimating the average income of a certain age group, we may use a method based on the assumption that we are sampling a normal population, whereas the population is highly skewed.

Indeed the most questions of robustness are difficult to answer. When it comes to questions of robustness, we face all sorts of difficulties mathematically and most of the parts can be resolved by computer simulations.

3.10 METHODS OF ESTIMATION

So far we have discussed the requisites of a good estimator. There may be many different estimators of one and the same parameter of a population. Hence it is desirable to have a general method that yield estimators with as many properties as possible. Now we will briefly outline some of the important methods for obtaining such estimators. The most commonly used methods are

(i) Method of moments.

(ii) Method of Maximum Likelihood Estimator (MLE).

(iii) Method of minimum variance.

(iv) Method of Least squares.

(v) Method of minimum Chi-square.

(vi) Method of inverse probability.

In this section we shall discuss about the first two methods only.

3.10.1 The method of moments

The method moments is one of the oldest methods among all the methods. The method of moments consists of equating the first few moments of a population to the corresponding moments of a sample, getting as many equations as are needed to solve for the unknown parameters of the population.

Definition:

The k^{th} sample moment of a set of observation $x_1, x_2, x_3 \ldots x_n$ is the mean of their k^{th} powers and it is denoted by m_k'.

(ie) $$m_k' = \frac{1}{n} \sum_{i=1}^{n} x_i^k.$$

Hence if a population has r parameters, the method of moments consists of solving the system of equations.

$$m_k' = \mu_k' \text{ where } k = 1, 2, 3 \ldots r \text{ for } r \text{ parameters}$$

WORKED EXAMPLES

Example: 1

Find the estimator of θ in the population with density function $f(x, \theta) = \theta x^{\theta - 1}$; $0 < x < 1$; $\theta > 0$, by the method of moments.

Solution:

The first moment about the origin of the population is given by

$$\mu_1' = \int_0^1 x f(x)\, dx = \int_0^1 x \cdot \theta \cdot x^{\theta - 1}\, dx$$

$$= \theta \int_0^1 x^{\theta}\, dx = \theta \left[\frac{x^{\theta+1}}{\theta+1} \right]_0^1 = \theta \left[\frac{1}{\theta+1} - 0 \right]$$

$$\therefore \mu_1' = \frac{\theta}{\theta+1}.$$

The first moment of the sample $(x_1, x_2, x_3 \ldots x_n)$ about the origin is given by

$$m_1' = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}.$$

By the method of moments we know that

$$\mu_k' = m_k' \quad \text{where} \quad k = 1, 2, 3 \ldots r$$

$$\therefore \qquad \mu_1' = m_1' \Rightarrow \bar{x} = \frac{\theta}{\theta + 1}$$

$$\Rightarrow \qquad \theta = \bar{x}(\theta + 1)$$

$$\Rightarrow \qquad \theta = \bar{x}\theta + \bar{x} \Rightarrow \theta - \bar{x}\theta = \bar{x}$$

$$\Rightarrow \qquad \theta[1 - \bar{x}] = \bar{x} \Rightarrow \theta = \frac{\bar{x}}{1 - \bar{x}}$$

Example: 2

Let $(x_1, x_2, x_3 \ldots x_n)$ be a random sample from the uniform population with the density function $f(x\,;\,a, b) = \frac{1}{b-a}$; $a < x < b$. Find the estimators of a and b by the method of moments.

Solution:

The first moment about the origin of the uniform population is given by

$$\mu_1' = \int_a^b x \cdot f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx$$

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

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

$$\therefore \quad \mu_1' = \frac{b+a}{2}.$$

The second moment about the origin of uniform population is given by

$$\mu_2' = \int_a^b x^2 \cdot f(x)\,dx = \int_a^b x^2 \cdot \left(\frac{1}{b-a}\right)dx$$

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

$$= \frac{1}{b-a}\frac{(b-a)}{3}[a^2+ab+b^2]$$

$$\therefore \qquad \mu_2' = \frac{1}{3}[a^2+ab+b^2].$$

The first moment of the sample $(x_1, x_2, x_3 \dots x_n)$ about the origin is given by

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

The second moment of the sample $(x_1, x_2, x_3 \dots x_n)$ about the origin is given by

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = s^2.$$

By the method of moments we know that $\mu_1' = m_1'$ and $\mu_2' = m_2'$.

$$\therefore \qquad \mu_1' = m_1' \Rightarrow \frac{b+a}{2} = \bar{x}$$

$$\Rightarrow \qquad a+b = 2\bar{x} \qquad \dots (1)$$

Similarly $\quad \mu_2' = m_2' \Rightarrow \frac{1}{3}(a^2+ab+b^2) = s^2$

$$\Rightarrow \qquad a^2+ab+b^2 = 3s^2 \qquad \dots (2)$$

Using the equation (1), we have $b = 2\bar{x} - a$

By substituting $b = 2\bar{x} - a$ in equation (2) we get

$$a^2 + a(2\bar{x}-a) + (2\bar{x}-a)^2 = 3s^2$$

$$\Rightarrow \quad a^2 + 2a\bar{x} - a^2 + 4\bar{x}^2 + a^2 - 4a\bar{x} - 3s^2 = 0$$

$$\therefore \qquad a^2 - 2a\bar{x} + 4\bar{x}^2 - 3s^2 = 0$$

$$\Rightarrow \qquad a^2 - (2\bar{x})a + (4\bar{x}^2 - 3s^2) = 0$$

This is a quadratic equation in terms of a.

$$\therefore\ a = \frac{2\bar{x} \pm \sqrt{(-2\bar{x})^2 - 4(4\bar{x}^2 - 3s^2)}}{2} \qquad \left(\because x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\right)$$

$$a = \frac{2\bar{x} \pm \sqrt{4\bar{x}^2 - 16\bar{x}^2 + 12s^2}}{2}$$

$$a = \bar{x} + \sqrt{\bar{x}^2 - 4\bar{x}^2 + 3s^2}$$

$$= \bar{x} \pm \sqrt{-3\bar{x}^2 + 3s^2}$$

$$\therefore\ a = \bar{x} \pm \sqrt{3(s^2 - \bar{x}^2)} \qquad \dots (3)$$

Similarly from equation (1); we have $a = 2\bar{x} - b$

By substituting $a = 2\bar{x} - b$ in equation (2), we get

$$b = \bar{x} \mp \sqrt{3(s^2 - \bar{x}^2)} \qquad \dots (4)$$

Since $a < b$, we have

$$a = \bar{x} - \sqrt{3(s^2 - \bar{x}^2)} \text{ and}$$

$$b = \bar{x} + \sqrt{3(s^2 - \bar{x}^2)}$$

Example: 3

Let $(x_1, x_2, x_3 \dots x_n)$ be a random sample from a population with density function $f(x\,;\,\theta, \mu) = \theta e^{-\theta(x-\mu)}\,;\ x > \mu$. Find the method of moments estimators of θ and μ.

Solution:

The first moment about the origin of the given population is

$$\mu_1' = \int_{\mu}^{\infty} x \cdot f(x)\,dx = \int_{\mu}^{\infty} x \cdot \theta e^{-\theta(x-\mu)}\,dx$$

$$\mu_1' = \theta e^{\mu\theta} \int_{\mu}^{\infty} x \cdot e^{-\theta x}\,dx$$

$$= \theta\, e^{\mu\theta} \left[x \left\{ \frac{e^{-\theta x}}{-\theta} \right\} - (1) \left\{ \frac{e^{-\theta x}}{(-\theta)^2} \right\} \right]_{\mu}^{\infty}$$

$$= \theta\, e^{\mu\theta} \left[-x \left(\frac{e^{-\theta x}}{\theta} \right) - \left(\frac{e^{-\theta x}}{\theta^2} \right) \right]_{\mu}^{\infty}$$

$$= -\theta\, e^{\mu\theta} \left[x \frac{e^{-\theta x}}{\theta} + \frac{e^{-\theta x}}{\theta^2} \right]_{\mu}^{\infty}$$

$$= -\theta\, e^{\mu\theta} \left[\{ 0 + 0 \} - \left\{ \mu \frac{e^{-\theta\mu}}{\theta} + \frac{e^{-\theta\mu}}{\theta^2} \right\} \right]$$

$$= \theta\, e^{\mu\theta} \left[\mu \frac{e^{-\theta\mu}}{\theta} + \frac{e^{-\theta\mu}}{\theta^2} \right]$$

$$= \theta\, e^{\mu\theta} \cdot \mu \cdot \frac{e^{-\mu\theta}}{\theta} + \theta\, e^{\mu\theta} \cdot \frac{e^{-\mu\theta}}{\theta^2}$$

$$= \mu + \frac{1}{\theta}$$

$$\therefore\ \mu_1' = \mu + \frac{1}{\theta} \qquad \text{... (1)}$$

$$\mu_2' = \int_{\mu}^{\infty} x^2 \cdot f(x)\, dx = \int_{\mu}^{\infty} x^2 \cdot \theta\, e^{-\theta x} \cdot e^{\mu\theta}\, dx$$

$$\mu_2' = \theta\, e^{\mu\theta} \int_{\mu}^{\infty} x^2 e^{-\theta x}\, dx$$

$$= \theta\, e^{\mu\theta} \left[(x^2) \left(\frac{e^{-\theta x}}{-\theta} \right) - (2x) \left(\frac{e^{-\theta x}}{\theta^2} \right) + (2) \left(\frac{e^{-\theta x}}{-\theta^3} \right) \right]_{\mu}^{\infty}$$

$$= \theta\, e^{\mu\theta} \left[-x^2 \frac{e^{-\theta x}}{\theta} - 2x \frac{e^{-\theta x}}{\theta^2} - 2 \frac{e^{-\theta x}}{\theta^3} \right]_{\mu}^{\infty}$$

$$= -\theta\, e^{\mu\theta} \left[x^2 \frac{e^{-\theta x}}{\theta} + 2x \frac{e^{-\theta x}}{\theta^2} + 2 \frac{e^{-\theta x}}{\theta^3} \right]_{\mu}^{\infty}$$

$$= -\theta\, e^{\mu\theta} \left[\{0+0+0\} - \left\{ \mu^2 \frac{e^{-\mu\theta}}{\theta} + 2\mu \frac{e^{-\mu\theta}}{\theta^2} + 2 \frac{e^{-\mu\theta}}{\theta^3} \right\} \right]$$

$$= -\theta\, e^{\mu\theta} \left[-\left\{ \mu^2 \frac{e^{-\mu\theta}}{\theta} + 2\mu \frac{e^{-\mu\theta}}{\theta^2} + 2 \frac{e^{-\mu\theta}}{\theta^3} \right\} \right]$$

$$= \theta\, e^{\mu\theta}\, e^{-\mu\theta} \left[\frac{\mu^2}{\theta} + \frac{2\mu}{\theta^2} + \frac{2}{\theta^3} \right]$$

$$= \theta \left[\frac{\mu^2}{\theta} + \frac{2\mu}{\theta^2} + \frac{2}{\theta^3} \right]$$

$$\mu_2' = \mu^2 + \frac{2\mu}{\theta} + \frac{2}{\theta^2} \qquad \dots (2)$$

The first moment of the sample is $\mu_1' = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$ ... (3)

The second moment of the sample is $\mu_2' = \frac{1}{n} \sum_{i=1}^{n} x_i^2 = s^2$ ... (4)

We know that by the method of moments we have

$$\mu_1' = m_1' \text{ and } \mu_2' = m_2'$$

$$\therefore\ \mu_1' = m_1' \Rightarrow \bar{x} = \mu + \frac{1}{\theta} \qquad \dots (5)$$

Also $\mu_2' = m_2' \Rightarrow \mu^2 + \frac{2\mu}{\theta} + \frac{2}{\theta^2} = s^2$. ... (6)

From equation (5), we have $\frac{1}{\theta} = \bar{x} - \mu$.

Substitute $\frac{1}{\theta} = \bar{x} - \mu$ in equation (6), we get

$$\mu^2 + 2\mu(\bar{x} - \mu) + 2(\bar{x} - \mu)^2 = s^2$$

$$\mu^2 + 2\mu\bar{x} - 2\mu^2 + 2(\bar{x}^2 + \mu^2 - 2\bar{x}\mu) - s^2 = 0$$

$$\mu^2 + 2\mu\bar{x} - 2\mu^2 + 2\bar{x}^2 + 2\mu^2 - 4\bar{x}\mu - s^2 = 0$$

$$\mu^2 - 2\bar{x}\mu + 2\bar{x}^2 - s^2 = 0$$

$$\Rightarrow \mu^2 - (2\bar{x})\mu + (2\bar{x}^2 - s^2) = 0$$

This is a quadratic equation in μ

$$\therefore \mu = \frac{2\bar{x} \pm \sqrt{(-2\bar{x})^2 - 4(1)(2\bar{x}^2 - s^2)}}{2}$$

$$\mu = 2\bar{x} \pm \frac{\sqrt{4\bar{x}^2 - 8\bar{x}^2 + 4s^2}}{2}$$

$$= \bar{x} \pm \sqrt{\bar{x}^2 - 2\bar{x}^2 + s^2}$$

$$\mu = \bar{x} \pm \sqrt{s^2 - \bar{x}^2}$$

$$\therefore \qquad \theta = \frac{1}{\bar{x} - \mu} \quad \text{(or)} \quad \theta = \frac{1}{\sqrt{s^2 - \bar{x}^2}}$$

$$\mu = \bar{x} - \sqrt{s^2 - \bar{x}^2}$$

Example: 4

For the probability mass function

$$f(x;p) = 3c_x \cdot \frac{p^x (1-p)^{3-x}}{1-(1-p)^3}; \; x = 1, 2, 3.$$

Obtain the estimator of p by the method of moments, if the frequencies at $x = 1, 2, 3$ respectively 22, 20, 18.

✍ **Solution:**

$$f(x,p) = \frac{1}{1-(1-p)^3} B(3;p)$$

The first moment about the origin is

$$\mu_1' = \frac{1}{1-(1-p)^3} \cdot 3p$$

The mean of the observed sample is given by

$$\bar{x} = \frac{1 \times 22 + 2 \times 20 + 3 \times 18}{22 + 20 + 18} = \frac{116}{60} \text{ (or) } \frac{29}{15}.$$

By the method of moments $\mu_1' = \bar{x}$.

$$\frac{3p}{3p - 3p^2 + p^3} = \frac{29}{15} \Rightarrow 29p^2 - 87p + 42 = 0$$

Solving this equation, we get

$$p = \frac{87 \pm 51.93}{58} = 2.395 \text{ (or) } 0.605.$$

Since 2.395 is inadmissible, $p = 0.605$.

Example: 5

A random variable X takes the values 0, 1, 2 with respective probabilities $\frac{1}{2} - \theta, \frac{\alpha}{2} + 2(1-\alpha)\theta$ and $\left(\frac{1-\alpha}{2}\right) + (2\alpha - 1)\theta$, where α and θ are the parameters. If a sample of size 75 drawn from the population yielded the values 0, 1, 2 with respective frequencies 27, 38, 10 respectively, find the estimators of α and θ by the method of moments.

Solution:

$$\mu_1' = E[X] = 0 \times \left(\frac{1}{2} - \theta\right) + 1 \times \left\{\frac{\alpha}{2} + 2(1-\alpha)\theta\right\} + 2 \times \left\{\frac{1-\alpha}{2} + (2\alpha - 1)\theta\right\}$$

$$= 1 - \frac{\alpha}{2} + 2\alpha\theta$$

$$\mu_2' = E[X^2] = 0^2 \times \left\{\frac{1}{2} - \theta\right\} + 1^2 \times \left\{\frac{\alpha}{2} + 2(1-\alpha)\theta\right\} + 2^2 \times \left\{\frac{1-\alpha}{2} + (2\alpha - 1)\theta\right\}$$

$$= 2 - \frac{3}{2}\alpha + (6\alpha - 2)\theta$$

$$m_1' = \frac{38 \times 1 + 10 \times 2}{75} = \frac{58}{75}$$

$$m_2' = s^2 = \frac{1}{75} [38 \times 1^2 + 10 \times 2^2] = \frac{78}{75}$$

By the method of moments, $\mu_1' = \bar{x}$ and $\mu_2' = s^2$

$$\Rightarrow \quad 1 - \frac{\alpha}{2} + 2\alpha\,\theta = \frac{58}{75} \text{ and}$$

$$2 - \frac{3}{2}\alpha + (6\alpha - 2)\,\theta = \frac{78}{75}$$

Solving the above two equations, we get

$$\alpha = \frac{34}{33} \text{ and } \theta = \frac{7}{50}$$

Example: 6

Given a random sample of size n from a gamma population, use the method of moments to obtain formulas for estimating the parameters α and β.

Solution:

We know that the r^{th} moment about the origin of the gamma distribution is

$$\mu_r' = \frac{\beta^r\,\Gamma(\alpha + r)}{\Gamma\alpha}.$$

The r^{th} moment about the origin of a random variable X, denoted by μ_r' is the expected value of X^r,

$$\therefore \; \mu_r' = E[X^r] = \sum_x x^r \cdot f(x) \quad \text{for} \quad r = 0, 1, 2 \ldots$$

when X is discrete and

$$\mu_r' = E[X^r] = \int_{-\infty}^{\infty} x^r \cdot f(x)\,dx \text{ when } X \text{ is continuous.}$$

The system of equations we have to solve is

$$m_1' = \mu_1' \text{ and } m_2' = \mu_2'.$$

$$m_1' = \mu_1' = \frac{\beta\,\Gamma\alpha+1}{\Gamma\alpha} = \frac{\beta\,\alpha\,\Gamma\alpha}{\Gamma\alpha} \qquad (\because \Gamma n+1 = n\,\Gamma n)$$

$$\Rightarrow \mu_1' = \alpha\,\beta$$

Then
$$\mu_2' = \frac{\beta^2\,\Gamma\alpha+2}{\Gamma\alpha} = \frac{\beta^2\,(\alpha+1)\,\Gamma\alpha+1}{\Gamma\alpha}$$

$$= \frac{\beta^2\,(\alpha+1)\,\alpha\,\Gamma\alpha}{\Gamma\alpha}$$

$$\mu_2' = \alpha\,\beta^2\,(\alpha+1)$$

$$\therefore\ m_1' = \mu_1' = \alpha\,\beta \text{ and } m_2' = \mu_2' = \alpha\,\beta^2\,(\alpha+1).$$

Solving for α and β, we get the following formulas for estimating the two parameters of gamma distribution.

$$\hat{\alpha} = \frac{(m_1')^2}{m_2' - (m_1')^2} \text{ and } \hat{\beta} = \frac{m_2' - (m_1')^2}{m_1'}$$

Since $m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ and $m_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, we can write

$$\hat{\alpha} = \frac{n\,\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \text{ and } \hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n\,\bar{x}}.$$

3.10.2 Method of Maximum Likelihood Estimation (MLE)

Prof. R.A. Fisher, the prominent statistician, proposed a general method of estimation called the method of maximum likelihood estimators (M.L.E). He had explained the advantages of this method by showing that it yields sufficient estimators whenever they exist and that maximum likelihood estimators are asymptotically minimum variance unbiased estimators.

Definition:

If $x_1, x_2, x_3 \ldots x_n$ are the values of a random sample from a population with the parameter θ, the likelihood function of the sample is given by $L[\theta] = f(x_1, x_2, x_3 \ldots x_n ; \theta)$ for values of θ within the given domain. Here $f(x_1, x_2, x_3, \ldots x_n ; \theta)$ is the value of the joint probability distribution or the joint probability density function of the random variables $X_1, X_2, X_3 \ldots X_n$ at $X_1 = x_1, X_2 = x_2 \ldots X_n = x_n$.

$$\text{(ie)} \quad L = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

Hence the method of maximum likelihood consists of maximizing the likelihood function with respect to θ, and we refer to the value of θ which maximize the likelihood function as the maximum likelihood estimate of θ.

Note:

The principle of maximum likelihood consists in finding an estimator of the parameter which maximizes L for variations in the parameter. Thus if there exists a function $\hat{\theta} = \hat{\theta}[x_1, x_2, x_3 \ldots x_n]$ of the sample values which maximizes L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is usually called Maximum Likelihood Estimator (M.L.E).

Thus $\hat{\theta}$ is the solution, if any of,

$$\frac{\partial L'}{\partial \theta} = 0 \text{ and } \frac{\partial^2 L}{\partial \theta^2} < 0.$$

Since $L > 0$, so is $\log L$ which shows that L and $\log L$ attain their extreme values (maxima or minima) at the same value of $\hat{\theta}$. The above two equations can be rewritten as

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0.$$

This equation is usually referred to as the Likelihood equation.

3.10.3 Properties of maximum likelihood estimators

Property 1: The first and second order derivatives $\frac{\partial \log L}{\partial \theta}$ and $\frac{\partial^2 \log L}{\partial \theta^2}$ exist and are continuous functions of θ in a range R, for almost all x.

For every θ in R

$$\frac{\partial}{\partial\theta}\log L < F_1(x) \text{ and } \left|\frac{\partial^2}{\partial\theta^2}\log L\right| < F_2(x) \text{ where } F_1(x) \text{ and } F_2(x)$$

are integrable functions over $(-\infty, \infty)$.

Property 2: The third order derivative $\frac{\partial^3}{\partial\theta^3}\log L$ exists such that $\left|\frac{\partial^3}{\partial\theta^3}\log L\right| < M(x)$ where $E[M(x)] < K$, a positive quantity.

Property 3: For every θ in R,

$$E\left[-\frac{\partial^2}{\partial\theta^2}\log L\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\left[-\frac{\partial^2}{\partial\theta^2}\log L\right] L\,dx_1, dx_2, \ldots dx_n$$

is finite and non-zero.

Property 4: The range of integration is independent of θ. But if the range of integration depends on θ, then $f(x, \theta)$ vanishes at the extremes depending on θ.

Theorem 5: Cramer Rao's theorem

With probability approaching unity as $n \to \infty$, the likelihood equation $\frac{\partial}{\partial\theta}\log L = 0$ has a solution which converges in probability to the true value θ_0.

(ie) M.L.E are consistent.

Theorem 6: Hazoor Bazar's theorem

Any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tending to unity as the sample size (n) tends to infinity.

Theorem 7: A consistent solution of the likelihood equation in asymptotically normally distributed about the true value θ_0. Thus $\hat{\theta}$ is asymptotically $N\left[\theta_0, \frac{1}{I(\theta_0)}\right]$.

The variance of M.L.E is defined by

$$\text{Var}[\hat{\theta}] = \frac{1}{I(\theta)} = \frac{1}{\left[E\left[-\frac{\partial^2}{\partial\theta^2}\log L\right]\right]}.$$

WORKED EXAMPLES

Example: 1

Given x successes in n trials, find the maximum likelihood e of the parameter θ of the corresponding binomial distribution.

Solution:

Since the likelihood function is

$$L[\theta] = f(x_1, x_2 \dots x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta).$$

We know that the probability mass function of binomial distr is

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

To find the value of θ which maximizes

$$L[\theta] = nC_x \cdot \theta^x (1-\theta)^{n-x}.$$

It will be convenient to make use of the value of θ which ma $L[\theta]$ will also maximize

$$\log [L[\theta]] = \log [nC_x \theta^x (1-\theta)^{n-x}]$$

$$\log [L[\theta]] = \log (nc_x) + \log \theta^x + \log (1-\theta)^{n-x}$$

$$\log [L[\theta]] = \log (nc_x) + x \log \theta + (n-x) \log (1-\theta)$$

Differentiating equation (1) partially with respect to θ on both

$$\frac{\partial}{\partial \theta} \log [L(\theta)] = 0 + x \cdot \frac{1}{\theta} + (n-x) \frac{1}{(1-\theta)} (-1).$$

$$\frac{\partial \log [L(\theta)]}{\partial \theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

We know that the condition for the maximum likelihood estin

$\frac{\partial \log L}{\partial \theta} = 0.$

$$\therefore \qquad \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$$

$$\frac{x}{\theta} = \frac{n-x}{1-\theta} \Rightarrow \frac{1-\theta}{\theta} = \frac{n-x}{x}$$

$$\frac{1}{\theta}-1=\frac{n-x}{x} \Rightarrow \frac{1}{\theta}=\frac{n-x}{x}+1$$

$$\frac{1}{\theta}=\frac{n-x+x}{x} \Rightarrow \frac{1}{\theta}=\frac{n}{x}$$

$$\therefore\ \theta=\frac{x}{n}$$

$\therefore$ We found that the likelihood function has a maximum at $\theta=\frac{x}{n}$. This is the maximum likelihood estimate of the binomial parameter θ and we refer to $\hat{\theta}=\frac{X}{n}$ as the corresponding maximum likelihood estimator.

Example: 2

If $x_1, x_2, x_3 \ldots x_n$ are the values of a random sample from an exponential population, find the maximum likelihood estimator of the parameter θ.

Solution:

Since the likelihood function is given by

$$L[\theta]=f(x_1, x_2, x_3 \ldots x_n ; \theta)=\prod_{i=1}^{n} f(x_i ; \theta).$$

We know that the probability density function of exponential distribution is $f(x)=\lambda e^{-\lambda x}$ where $x \geq 0$.

The mean of the exponential distribution is $E[X]=\overline{X}=\frac{1}{\lambda}$

The variance of the exponential distribution is $\sigma^2=\frac{1}{\lambda^2}$.

$\therefore$ For the parameter θ ; $f(x ; \theta)=\left(\frac{1}{\theta}\right) e^{\frac{-1}{\theta}(x)}$

$\therefore$ From the above condition; $f(x_i\,;\theta) = \left(\dfrac{1}{\theta}\right)^n e^{\frac{-1}{\theta}\left(\sum\limits_{i=1}^{n} x_i\right)}$.

$$\therefore \quad L[\theta] = \left(\frac{1}{\theta}\right)^n e^{\frac{-1}{\theta}\left[\sum\limits_{i=1}^{n} x_i\right]}.$$

Now take log on both sides.

$$\log[L(\theta)] = \log\left[\left(\frac{1}{\theta}\right)^n e^{\frac{-1}{\theta}\left[\sum\limits_{i=1}^{n} x_i\right]}\right]$$

$$\log[L(\theta)] = \log\left(\frac{1}{\theta}\right)^n + \log e^{\frac{-1}{\theta}\left[\sum\limits_{i=1}^{n} x_i\right]}$$

$$= n\log\left(\frac{1}{\theta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i \log e$$

$$\log[L(\theta)] = n\log\left(\frac{1}{\theta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i \qquad \text{... (1)}$$

$$(\because \log_e e = 1)$$

Differentiate (1) partially with respect to θ, on both sides.

$$\frac{\partial L(\theta)}{\partial\theta} = n\,\frac{1}{\left(\frac{1}{\theta}\right)}\cdot\left(\frac{1}{-\theta^2}\right) + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

$$= n\cdot\theta\left(-\frac{1}{\theta^2}\right) + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

$$\frac{\partial L(\theta)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

$$\therefore \frac{\partial L(\theta)}{\partial\theta} = 0 \Rightarrow -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0$$

$$\frac{n}{\theta} = \frac{1}{\theta^2} \sum_{i=1}^{n} x_i$$

$$\frac{\theta^2}{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i \Rightarrow \theta = \bar{x}$$

$\therefore$ The maximum likelihood estimator is $\hat{\theta} = \bar{x}$.

Example: 3

If $x_1, x_2, x_3 \ldots x_n$ are the values of a random sample of size n from a uniform population with $\alpha = 0$, find the maximum likelihood estimator of β.

Solution:

We know that the probability density function of uniform distribution is

$$f(x) = \frac{1}{\beta - \alpha}; \ \alpha < x < \beta.$$

The mean of the uniform distribution is $E[X] = \bar{X} = \dfrac{\beta + \alpha}{2}$.

The variance of uniform distribution is $\sigma^2 = \dfrac{1}{12}(\beta - \alpha)^2$.

Since the likelihood function is given by

$$L[\theta] = f(x_1, x_2, x_3 \ldots x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

For the parameter β (i.e. $\alpha = 0$)

$$f(x; \beta) = \frac{1}{\beta} \text{ since } \alpha = 0$$

$$\therefore f(x_i; \beta) = \left(\frac{1}{\beta}\right)^n$$

$$L[\beta] = \prod_{i=1}^{n} f(x_i; \beta) = \left(\frac{1}{\beta}\right)^n.$$

Now take log on both sides. We get

$$\log [L(\beta)] = \log \left[\frac{1}{\beta}\right]^n$$

$$\log [L(\beta)] = n \log \left(\frac{1}{\beta}\right). \qquad \ldots (1)$$

Differentiate equation (1) partially with respect to β on both sides

$$\frac{\partial [\log L(\beta)]}{\partial \beta} = n \cdot \frac{1}{\left(\frac{1}{\beta}\right)} \cdot \left(-\frac{1}{\beta^2}\right)$$

$$\Rightarrow \frac{\partial}{\partial \beta} [\log L(\beta)] = -\frac{n\beta}{\beta^2} = -\frac{n}{\beta}.$$

Since $\quad \dfrac{\partial}{\partial \beta} [\log [L(\theta)] = 0 \Rightarrow -\dfrac{n}{\beta} = 0$

$$\Rightarrow \frac{n}{\beta} = 0.$$

For β greater than or equal to the largest of the $x's$ and 0 otherwise. Since the value of this likelihood function increases as β decreases, we must take β as small as possible and it follows that the maximum likelihood estimator of β is Y_n, the n^{th} order statistic.

Example: 4

If $X_1, X_2, X_3 \ldots X_n$ constitute a random sample of size n from a normal population with mean μ and the variance σ^2, find joint maximum likelihood estimates of these two parameters.

Solution:

Since likelihood function is given by

$$L[\theta] = f(x_1, x_2, x_3 \ldots x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

$$\Rightarrow \quad L[\mu, \sigma^2] = \prod_{i=1}^{n} n(x_i ; \mu, \sigma)$$

We know that the probability density function of normal distribution is

$$f(x\,;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\text{ where }-\infty<x<\infty,\ -\infty<\mu<\infty,\ \sigma>0.$$

The mean of the normal distribution is $E[X]=\mu$.

The variance of normal distribution is $\text{Var}[X]=\sigma^2$.

$\therefore$ The S.D of normal distribution is σ.

$$\text{Here } L[\mu,\sigma^2]=\prod_{i=1}^{n}\ n(x_i\,;\mu,\sigma)$$

$$\therefore\ f(x_i\,;\mu,\sigma)=\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n\cdot e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\therefore\quad L[\mu,\sigma^2]=\left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n\cdot e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Now take log on both sides.

$$\log L[\mu,\sigma^2]=\log\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\cdot e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\right]$$

$$\log L[\mu,\sigma^2]=\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n+\log\left\{e^{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\right\}$$

$$\log L[\mu,\sigma^2]=-\frac{n}{2}\log(\sigma^2)-\frac{n}{2}\log(2\pi)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \qquad \dots(1)$$

The likelihood equations for the simultaneous estimations of μ and σ^2 are $\frac{\partial}{\partial\mu}\log L=0$ and $\frac{\partial}{\partial\sigma^2}\log L=0$.

Differentiate (1) partially with respect to μ on both sides

$$\frac{\partial}{\partial\mu}[\log L(\mu,\sigma^2)]=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\ 2(x_i-\mu)\ (-1)$$

$$\frac{\partial}{\partial \mu}[L(\mu, \sigma^2)] = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$$

Since $\frac{\partial}{\partial \mu}[\log L(\mu, \sigma^2)] = 0$

$$\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0$$

$$\Rightarrow n\mu = \sum_{i=1}^{n} x_i \Rightarrow \mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\Rightarrow \mu = \bar{x} \qquad \text{... (2)}$$

Hence $\mu = \bar{x}$ is the maximum likelihood estimator for μ is the sample mean.

Now differentiate (1) partially with respect to σ^2, we get

$$\frac{\partial}{\partial \sigma^2}\{\log L(\mu, \sigma^2)\} = -\frac{n}{2}\left(\frac{1}{\sigma^2}\right) + \frac{1}{2}\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Since $\frac{\partial}{\partial \sigma^2}[\log L(\mu, \sigma^2)] = 0$,

$$-\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

$$\frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = \frac{n}{2}\frac{1}{\sigma^2}$$

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 = n$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \sigma^2$$

$$\therefore \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \qquad \dots (3)$$

Hence from equations (2) and (3)

We have $\mu = \overline{x}$ and $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$

$\therefore \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 = s^2$ is the sample variance.

Here it is noted that $E(\mu) = E[\overline{x}] = \mu$ and

$$E[\sigma^2] = E[s^2] \neq \sigma^2$$

Hence, the maximum likelihood estimators need not necessarily by unbiased.

Example: 5

Find the maximum likelihood estimator for the parameter λ of a Poisson distribution from n sample values. Also find its variance.

Solution:

Since the likelihood function is given by

$$L[\theta] = f(x_1, x_2, x_3 \dots x_n ; \theta] = \prod_{i=1}^{n} f(x_i ; \theta).$$

We know that the probability mass function of the Poisson distribution with parameter λ is given by

$$P[X=x]=p(x,\lambda)=\frac{e^{-\lambda}\cdot\lambda^{x}}{x!},\ 0\le x<\infty$$

$$\therefore\ L[\lambda]=\prod_{i=1}^{n} f(x_i;\lambda)=\frac{e^{-n\lambda}\ \lambda^{\sum_{i=1}^{n} x_i}}{x_1!\,x_2!\ldots x_n!}$$

By taking log on both sides, we get

$$\log L[\lambda]=\log\left[\frac{e^{-n\lambda}\cdot\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\,x_2!\ldots x_n!}\right]$$

$$\log L[\lambda]=\log(e^{-n\lambda})+\log\left(\lambda^{\sum_{i=1}^{n} x_i}\right)-\log[x_1!\,x_2!\ldots x_n!]$$

$$\log L[\lambda]=-n\lambda+\left(\sum_{i=1}^{n} x_i\right)\log\lambda-\sum_{i=1}^{n}\log(x_i!)$$

$$\log L(\lambda)=-n\lambda+n\bar{x}\log\lambda-\sum_{i=1}^{n}\log(x_i!) \qquad \ldots (1)$$

Differentiate (1), partially w.r.to λ on both sides

$$\frac{\partial}{\partial\lambda}[\log L(\lambda)]=-n+n\bar{x}\cdot\frac{1}{\lambda}$$

Since $\frac{\partial}{\partial\lambda}[\log L(\lambda)]=0$ then

$$-n+\frac{n\bar{x}}{\lambda}=0$$

$$\Rightarrow n=\frac{n\bar{x}}{\lambda}\Rightarrow\lambda=\frac{n\bar{x}}{n}=\bar{x}$$

$$\therefore\ \lambda=\bar{x}$$

Thus the maximum likelihood estimator for λ is the sample mean $\bar{x}$.

The variance of the estimate is given by

$$\frac{1}{\text{Var}[\lambda]} = E\left[-\left(\frac{\partial^2}{\partial \lambda^2} \log L(\lambda)\right)\right]$$

$$= E\left[-\frac{\partial}{\partial \lambda}\left(-n + \frac{n\bar{x}}{\lambda}\right)\right]$$

$$= E\left[-\left(-\frac{n\bar{x}}{\lambda^2}\right)\right]$$

$$= E\left[\frac{n\bar{x}}{\lambda^2}\right]$$

$$= \frac{n}{\lambda^2} E[\bar{x}]$$

$$= \frac{n}{\lambda^2}(\lambda)$$

$$\frac{1}{\text{Var}[\lambda]} = \frac{n}{\lambda}$$

$$\therefore \qquad \text{Var}[\lambda] = \frac{\lambda}{n}.$$

Example: 6

Find the maximum likelihood estimator of the parameters α and λ (λ being large) of the distribution

$$f(x;\alpha,\lambda) = \frac{1}{\Gamma\lambda}\left(\frac{\lambda}{\alpha}\right)^{\lambda} e^{\frac{-\lambda x}{\alpha}} x^{\lambda-1};\ 0 \le x < \infty, \lambda > 0$$

you may use that for large values of λ,

$$\psi(\lambda) = \frac{\partial}{\partial \lambda}\log \Gamma\lambda = \log \lambda - \frac{1}{2\lambda} \text{ and } \psi'(\lambda) = \frac{1}{\lambda} + \frac{1}{2\lambda^2}.$$

Solution:

Let $x_1, x_2, x_3 \ldots x_n$ be a random sample of size n from the given population.

Since the likelihood function is given by

$$L = f(x_1, x_2, x_3 \ldots x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta).$$

Then for this problem

$$L = \prod_{i=1}^{n} f(x_i ; \alpha, \lambda)$$

$$= \left(\frac{1}{\Gamma\lambda}\right)^n \left(\frac{\lambda}{\alpha}\right)^{n\lambda} e^{-\frac{\lambda}{\alpha} \sum_{i=1}^{n} x_i} \sum_{i=1}^{n} (x_i^{\lambda-1}).$$

By taking log on both sides, we get

$$\log L[x ; \alpha, \lambda] = \log\left[\frac{1}{\Gamma\lambda}\right]^n + \log\left[\frac{\lambda}{\alpha}\right]^{n\lambda} + \log e^{-\frac{\lambda}{\alpha} \sum_{i=1}^{n} x_i}$$

$$+ \log\left[\sum_{i=1}^{n} (x_i^{\lambda-1})\right].$$

$$\log L = n \log\left[\frac{1}{[\Gamma\lambda]}\right] + n\lambda \log\left[\frac{\lambda}{\alpha}\right] - \frac{\lambda}{\alpha} \sum_{i=1}^{n} x_i$$

$$+ (\lambda - 1) \sum_{i=1}^{n} \log(x_i).$$

$$\therefore \log L = n[\log(1) - \log[\Gamma\lambda] + n\lambda[\log\lambda - \log\alpha]$$

$$-\frac{\lambda}{\alpha} \sum_{i=1}^{n} x_i + (\lambda - 1) \sum_{i=1}^{n} \log(x_i).$$

$$\log L = -n \log(\Gamma\lambda) + n\lambda[\log\lambda - \log\alpha] - \frac{\lambda}{\alpha} \sum_{i=1}^{n}$$

$$+ (\lambda - 1) \sum_{i=1}^{n} \log(x_i).$$

If G is the Geometric mean of $x_1, x_2, x_3 \ldots x_n$, then

$$\log G = \frac{1}{n} \sum_{i=1}^{n} \log (x_i) \Rightarrow n \log G = \sum_{i=1}^{n} \log (x_i) .$$

$$\therefore \log L = -n \log (\Gamma \lambda) + n\lambda [\log \lambda - \log \alpha]$$

$$-\frac{\lambda}{\alpha} \cdot n\bar{x} + (\lambda - 1)\, n \log G \qquad \text{... (1)}$$

where G is independent of λ and α.

The likelihood equations for the simultaneous estimation of α and λ are

$$\frac{\partial}{\partial \alpha} \log L = 0 \quad \text{and} \quad \frac{\partial}{\partial \lambda} \log L = 0$$

Now, from (1) we get $\frac{\partial}{\partial \alpha} \log L = n\lambda \left[-\frac{1}{\alpha} \right] + \frac{\lambda}{\alpha^2} n\bar{x} = 0$

$$\frac{-n\lambda}{\alpha} + \frac{\lambda}{\alpha^2} + n\bar{x} = 0$$

$$\frac{\lambda}{\alpha^2} n\bar{x} = \frac{n\lambda}{\alpha}$$

$$\frac{\alpha}{\alpha^2} = \frac{n\lambda}{\lambda n \bar{x}}$$

$$\Rightarrow \qquad \frac{1}{\alpha} = \frac{1}{\bar{x}} \Rightarrow \alpha = \bar{x}.$$

Also $\qquad \frac{\partial}{\partial \lambda} \log L = 0$

$$-n\left[\log \lambda - \frac{1}{2\lambda} \right] + n\left[1 \cdot (\log \lambda - \log \alpha) + \lambda \cdot \frac{1}{\lambda} \right] - \frac{n\bar{x}}{\alpha} + n \log G = 0$$

$$\Rightarrow \qquad \frac{1}{2\lambda} + \left[1 - \log \alpha + \log G - \frac{\bar{x}}{\alpha} \right] = 0$$

$$\Rightarrow \qquad 1 + 2\lambda [\log G - \log \bar{x}] = 0$$

$$\Rightarrow \qquad 1 - 2\lambda \log \left[\frac{\bar{x}}{G} \right] = 0$$

$$\Rightarrow \qquad 2\lambda \log \left[\frac{\bar{x}}{G} \right] = 1$$

$$\Rightarrow \qquad \lambda = \frac{1}{2 \log\left(\frac{\bar{x}}{G}\right)}.$$

$\therefore$ Hence the maximum likelihood estimators for α and λ are given by

$$\alpha = \bar{x} \text{ and } \lambda = \frac{1}{2 \log\left(\frac{\bar{x}}{G}\right)}.$$

Example: 7

A random sample X has a distribution with the density function

$$f(x) = \begin{cases} (\alpha+1)\, x^{\alpha}\,; \\ \qquad 0\,; & \text{otherwise} \end{cases}$$

and a random sample of size 8 produces the data 0.2, 0.4, 0.8, 0.5, 0.7, 0.9, 0.8, 0.9.

Find the maximum likelihood estimate of the unknown parameter α, it is given that

$$\ln[0.0145152] = -4.2326.$$

Solution:

Let us choose a random sample $X_1, X_2 \ldots X_n$ of size n from the population of X. Since the maximum likelihood estimator is given by

$$L = f(x_1, x_2, x_3 \ldots x_n\,;\theta) = \prod_{i=1}^{n} f(x_i\,;\theta).$$

For this problem, the above equation is rewritten as

$$L[\alpha] = \prod_{i=1}^{n} f(x_i\,;\alpha) = \prod_{i=1}^{n} (\alpha+1)\, x^{\alpha}.$$

$$\therefore \qquad L[\alpha] = (\alpha+1)^n \sum_{i=1}^{n} x_i^{\alpha}.$$

Take log on both sides, then we get

$$\log L[\alpha] = \log\left[(\alpha+1)^n \sum_{i=1}^{n} x_i^{\alpha}\right]$$

$$= \log(\alpha+1)^n + \log\left(\sum_{i=1}^{n} x_i^{\alpha}\right)$$

$$\log L[\alpha] = n\log(\alpha+1) + \alpha\log\left[\sum_{i=1}^{n} x_i\right]$$

$$\Rightarrow \log L[\alpha] = n\log(\alpha+1) + \alpha\log[x_1+x_2+x_3+\ldots+x_n] \qquad \ldots (1)$$

The condition for the maximum likelihood estimator is

$$\frac{\partial}{\partial\alpha}[\log[L(\alpha)]] = 0.$$

Now differentiate (1) partially with respect to α, then we get

$$\frac{\partial}{\partial\alpha}\log L[\alpha] = n\left(\frac{1}{\alpha+1}\right) + \log[x_1+x_2+x_3+\ldots+x_n] = 0$$

$$\therefore \log[x_1+x_2+x_3+\ldots+x_n] = -\frac{n}{\alpha+1}.$$

For the given sample, we have

$$\log[0.2\times0.4\times0.8\times0.5\times0.7\times0.9\times0.8\times0.9] = -\frac{8}{\alpha+1}$$

$$\Rightarrow \log[0.0145152] = -\frac{8}{\alpha+1}$$

$$-4.2326 = -\frac{8}{\alpha+1}$$

$$\therefore \qquad \alpha+1 = \frac{8}{4.2326}$$

$$\Rightarrow \qquad \alpha = \frac{8}{4.2326} - 1$$

$$\therefore \qquad \alpha = 0.8901$$

$\therefore$ The maximum likelihood estimator for α is 0.8901.

Example: 8

The pdf of a random variable X is assumed to be of the form $f(x) = cx^{\alpha}$; $0 \le x \le 1$ for some number and constant c. If $X_1, X_2, X_3 \ldots X_n$ is a random sample of size n, then find the maximum likelihood estimator of α.

Solution:

Since the maximum likelihood estimator is given by

$$L = f(x_1, x_2, x_3 \ldots x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta) .$$

Before finding the M.L.E we have to find the value of the constant c.

We know that the total pdf is $\int f(x)\, dx = 1$

For this problem we have $\int_0^1 cx^{\alpha}\, dx = 1$

$$\Rightarrow \qquad c\left[\frac{x^{\alpha+1}}{\alpha+1}\right]_0^1 = 1$$

$$\Rightarrow \qquad c\left[\frac{1}{\alpha+1} - 0\right] = 1$$

$$\Rightarrow \qquad c = \alpha + 1$$

∴ The function given in this problem is $f(x) = (\alpha + 1)\, x^{\alpha}$.

$$\therefore \quad L[\alpha] = \prod_{i=1}^{n} f(x_i ; \alpha) = \prod_{i=1}^{n} (\alpha + 1)\, x^{\alpha} .$$

$$\therefore \quad L[\alpha] = (\alpha + 1)^n \sum_{i=1}^{n} (x_i^{\alpha}) .$$

Take log on bothsides, then we get

$$\log L[\alpha] = \log\left[(\alpha + 1)^n \sum_{i=1}^{n} (x_i^{\alpha})\right]$$

$$= \log [(\alpha+1)^n] + \log \left[\sum_{i=1}^{n} (xi)^{\alpha} \right]$$

$$= n \log (\alpha+1) + \alpha \log \left[\sum_{i=1}^{n} x_i \right]$$

$$\log L[\alpha] = n \log (\alpha+1) + \alpha \log [x_1 + x_2 + x_3 \ldots + x_n] \qquad \ldots (1)$$

The condition for maximum likelihood estimator is

$$\frac{\partial}{\partial \alpha} [\log L(\alpha)] = 0 .$$

Now differentiate (1) partially with respect to α, then

$$\frac{\partial}{\partial \alpha} [\log L(\alpha)] = n \cdot \frac{1}{(\alpha+1)} + \log [x_1 + x_2 + x_3 \ldots + x_n] = 0$$

$$\therefore \frac{n}{\alpha+1} = -\log [x_1 + x_2 \ldots + x_n]$$

$$\Rightarrow -\frac{n}{\log [x_1 + x_2 + \ldots + x_n]} = \alpha + 1$$

$$\therefore \alpha = -1 - \frac{n}{\log [x_1 + x_2 + \ldots + x_n]}$$

which is the maximum likelihood estimator for α.

Example: 9

Find the maximum likelihood estimator of the parameter λ of the Weibull distribution $f(x) = \lambda \alpha x^{\alpha-1} e^{-\lambda x^{\alpha}}$ for $n > 0$, using a sample of size n, assuming that α is known.

Solution:

Let $X_1, X_2 \ldots X_n$ be a random sample of size n. The maximum likelihood function is

$$L[\lambda] = \prod_{i=1}^{n} f(x_i ; \lambda) = \prod_{i=1}^{n} \lambda \alpha x_i^{\alpha-1} e^{-\lambda x_i^{\alpha}} ; \; x > 0.$$

$$L(\lambda) = \lambda^n \alpha^n \sum_{i=1}^{n} x_i^{\alpha - 1} \cdot e^{-\lambda \sum_{i=1}^{n} x_i^{\alpha}}.$$

Take log on both sides. Then we get

$$\log [L(\lambda)] = \log(\lambda^n) + \log(\alpha^n) + \log \sum_{i=1}^{n} x_i^{\alpha - 1} + \log e^{-\lambda \sum_{i=1}^{n} x_i^{\alpha}}$$

$$\log L(\lambda) = n \log \lambda + n \log \alpha + (\alpha - 1) \sum_{i=1}^{n} \log(x_i) - \lambda \sum_{i=1}^{n} x_i^{\alpha} \quad \ldots ($$

The maximum likelihood estimate equation is

$$\frac{\partial}{\partial \lambda} \log [L(\lambda)] = 0$$

Differentiating (1) partially with respect to λ, we get

$$\frac{\partial}{\partial \lambda} \log [L(\lambda)] = n \cdot \frac{1}{\lambda} - \sum_{i=1}^{n} x_i^{\alpha} = 0.$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} x_i^{\alpha} \Rightarrow \lambda = \frac{n}{\sum_{i=1}^{n} x_i^{\alpha}}$$

This is the maximum likelihood estimator of λ.

Example: 10

The lifetime of a device has a pdf

$$f(x) = 3a^3 x^{-4} \text{ where } x \geq a.$$

For a random sample of size n, find the maximum likelihood estimat of the parameter a.

Solution:

Let us choose a random sample $X_1, X_2, X_3 \ldots X_n$ of size n from t population. The maximum likelihood estimator is given by

$$L[a] = f(x_1, x_2, x_3 \ldots, x_n ; a) = \prod_{i=1}^{n} f(x_i ; a).$$

$$\therefore \; L(a) = \prod_{i=1}^{n} [3a^3 x_i^{-4}] = 3^n a^{3n} \sum_{i=1}^{n} (x_i^{-4}).$$

Take log on both sides, we get

$$\log [L(a)] = \log \left[3^n a^{3n} \sum_{i=1}^{n} x_i^{-4} \right]$$

$$\log [L(a)] = \log (3^n) + \log (a^{3n}) + \log \left(\sum_{i=1}^{n} x_i^{-4} \right)$$

$$\log [L(a)] = n \log 3 + 3n \log a - 4 \left[\sum_{i=1}^{n} \log x_i \right] \quad \ldots (1)$$

The condition for M.L.E is $\frac{\partial}{\partial a} \log [L(a)] = 0$.

$$\therefore \frac{\partial}{\partial a} \log [L(a)] = 3n \cdot \frac{1}{a} = 0.$$

$$\Rightarrow \quad \frac{3n}{a} = 0.$$

Which does not yield a solution. So we choose a, such that $L(a)$ is maximum, which occurs if $a =$ min $[X_1, X_2, X_3 \ldots X_n]$

Hence the maximum likelihood estimator of a is

$$a = \min [X_1, X_2, X_3 \ldots X_n].$$

Example: 11

A sample of n independent observations is drawn from the rectangular population

$$f(x, \beta) = \begin{cases} \dfrac{1}{\beta}; & 0 < x \leq \beta, 0 < \beta < \infty \\ 0; & \text{otherwise} \end{cases}$$

Find the maximum likelihood estimator for β.

✍ **Solution:**

Since the likelihood estimator function is given by

$$L = f(x_1, x_2, x_3 \ldots x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Then for this problem

$$L = \prod_{i=1}^{n} f(x_i; \beta) = \left(\frac{1}{\beta}\right)^n$$

$$\therefore \qquad L[\beta] = \left(\frac{1}{\beta}\right)^n \qquad \text{[Refer Example 3]}$$

Now take log on both sides, then we get

$$\log [L(\beta)] = \log \left[\frac{1}{\beta}\right]^n$$

$$\log [L(\beta)] = n \log \left[\frac{1}{\beta}\right] \qquad \ldots (1)$$

Differentiate equation (1) partially with respect to β on both sides

$$\frac{\partial}{\partial \beta} \log [L(\beta)] = n \cdot \frac{1}{\left(\frac{1}{\beta}\right)} \cdot \left(-\frac{1}{\beta^2}\right)$$

$$= \frac{-n\beta}{\beta^2}$$

$$= -\frac{n}{\beta}$$

Since $\frac{\partial}{\partial \beta}[\log L(\beta)] = 0 \Rightarrow -\frac{n}{\beta} = 0 \Rightarrow \frac{n}{\beta} = 0$.

$\therefore$ Here $\beta = \infty$, an obviously absurd result.

So we have to shows β so that $L[\beta]$ is maximum. Now L is maximum if β is minimum. Let $x_1, x_2, x_3 \dots x_n$ be the ordered sample of n independent observations from the given population so that,

$$0 \leq x_1 \leq x_2 \leq x_3 \dots \leq x_n \leq \beta \Rightarrow \beta \geq x_n.$$

Hence the minimum value of β is consistent with the sample is x_n, the largest sample observations $\beta = x_n$.

Example: 12

Obtain the maximum likelihood estimators for α and β for the rectangular population

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

Solution:

Since the likelihood function is given by

$$L[\theta] = f(x_1, x_2, x_3 \dots x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

We know that the probability density function of uniform or rectangular distribution is given by

$$f(x) = \frac{1}{\beta - \alpha}; \ \alpha < x < \beta.$$

$$L[\alpha, \beta] = f(x_1, x_2, x_3 \dots x_n; \alpha, \beta) = \prod_{i=1}^{n} \left(\frac{1}{\beta - \alpha}\right)$$

$\therefore$ In this case $L[\alpha, \beta] = \left(\frac{1}{\beta - \alpha}\right)^n$.

Now take log on both sides

$$\log \{ L [\alpha, \beta] \} = \log \left\{ \left(\frac{1}{\beta - \alpha} \right)^n \right\}$$

$$\log [L (\alpha, \beta)] = n \cdot \log \left(\frac{1}{\beta - \alpha} \right) = - n \log (\beta - \alpha) \quad ...(1)$$

Now differentiate (1) partially with respect to α and β.

$$\therefore \frac{\partial}{\partial \alpha} \log [L (\alpha, \beta)] = - n \cdot \frac{1}{(\beta - \alpha)} (-1)$$

$$\therefore \qquad \frac{\partial}{\partial \alpha} \log [L (\alpha, \beta)] = \frac{n}{\beta - \alpha}$$

Since $\quad \dfrac{\partial}{\partial \alpha} \log [L (\alpha, \beta)] = 0 \Rightarrow \dfrac{n}{\beta - \alpha} = 0.$

$\therefore$ $\beta - \alpha = \infty$ which is an obviously negative result.

Now $\quad \dfrac{\partial}{\partial \beta} \log [L (\alpha, \beta)] = - n \cdot \dfrac{1}{(\beta - \alpha)} (1) = - \dfrac{n}{\beta - \alpha}$

Since $\quad \dfrac{\partial}{\partial \beta} \log [L (\alpha, \beta)] = 0 \Rightarrow - \dfrac{n}{\beta - \alpha} = 0 \Rightarrow \dfrac{n}{\beta - \alpha} = 0$

Again in this ease also $\beta - \alpha = \infty$ which is an obvious negative.

So we find M.L.E's for α and β by another form.

Now $L [\alpha, \beta] = \left(\dfrac{1}{\beta - \alpha} \right)^n$ is maximum if $(\beta - \alpha)$ is minimum. β takes the minimum possible value and α takes the maximum possible value. Hence as in Example (7),

$\alpha \le x_1 \le x_2 \le x_3 \ldots \le x_n \le \beta$. Thus $\beta > x_n$ and $\alpha \le x_n$. Hence the minimum possible value of β consistent with the sample is x_n and the maximum possible value of α consistent with the sample is x_1.

Hence L is maximum if $\beta = x_n$ and $\alpha = x_1$.

$\alpha = x_1 =$ Smallest sample observation and

$\beta = x_n =$ Largest sample observation.

Example: 13

Obtain the maximum likelihood estimators of α and β for a random sample from the exponential population.

$$f(x\,;\alpha,\beta)=y_0\,e^{-\beta(x-\alpha)},\ \alpha\le x\le\infty.$$

y_0 being a constant.

Solution:

Let us determine the constant y_0 from the total area under a probability curve is unity.

$$\therefore\quad y_0\int_{\alpha}^{\infty} e^{-\beta(x-\alpha)}\,dx=1$$

$$y_0\left[\frac{e^{-\beta(x-\alpha)}}{-\beta}\right]_{\alpha}^{\infty}=1$$

$$\Rightarrow\quad \frac{-y_0}{\beta}\,[e^{-\infty}-e^{-0}]=1$$

$$\Rightarrow\quad -\frac{y_0}{\beta}\,[0-1]=1$$

$$\Rightarrow\quad \frac{y_0}{\beta}=1$$

$$\therefore\quad y_0=\beta$$

Then the given probability density function becomes

$$\therefore\ f(x\,;\ \alpha,\beta)=\beta\,e^{-\beta(x-\alpha)},\ \alpha\le x\le\infty.$$

If x_1, x_2, x_3, ..., x_n is a random sample of n observations, from this population, then

$$L=\prod_{i=1}^{n} f(x_i\,;\alpha,\beta)=\prod_{i=-1}^{n}\beta\,e^{-\beta(x-\alpha)}$$

$$L=\beta^n\left[e^{-\beta\sum\limits_{i=1}^{n}(x_i-\alpha)}\right]=\beta^n\cdot e^{-n\beta(\bar{x}-\alpha)}$$

Now take log on both sides, then we get

$$\log [L] = \log [\beta^n \cdot e^{-n\beta(\bar{x}-\alpha)}]$$

$$= \log (\beta)^n + \log [e^{-n\beta(\bar{x}-\alpha)}]$$

$$\log L = n \log B - n\beta(\bar{x}-\alpha) \qquad \text{... (1)}$$

The likelihood equations for estimating α and β are given by

$$\frac{\partial}{\partial\alpha}\log L = 0 \quad \text{and} \quad \frac{\partial}{\partial\beta}\log L = 0.$$

Differentiate equation (1) partially with respect to α and β.

$$\frac{\partial}{\partial\alpha}\log L = -n\beta(-1) = 0$$

$$\Rightarrow n\beta = 0 \qquad \text{... (2)}$$

$$\Rightarrow \beta = 0$$

$$\frac{\partial}{\partial\beta}\log L = n \cdot \frac{1}{\beta} - n(\bar{x}-\alpha) = 0$$

$$\Rightarrow \frac{n}{\beta} - n(\bar{x}-\alpha) = 0\,. \qquad \text{... (3)}$$

Substitute $(\beta = 0)$ equation (2) in equation (3), we get

$\alpha = \infty$ which is a absurd result.

(ie) $\beta = 0$ and $\alpha = \infty$ are inadmissible values.

Thus the likelihood equations fail to give valid estimates of α and β by maximizing L.

L is maximum $\Rightarrow \log L$ is maximum.

From equation (1), $\log L$ is maximum for any value of β, if $(\bar{x}-\alpha)$ is minimum which is so if α is maximum.

If $x_1, x_2, x_3 \ldots x_n$ is ordered sample then

$$\alpha \le x_1 \le x_2 \le x_3 \ldots \le x_n \le \infty.$$

So that the maximum value of α consistent with the sample is x_1, the smallest sample observation $\alpha = x_1$.

Consequently, from equation (3) we have

$$\frac{1}{\beta} = (\bar{x} - \alpha) = \bar{x} - x_1 \Rightarrow \beta = \frac{1}{\bar{x} - x_1}.$$

Hence maximum likelihood estimators for α and β are given by

$$\alpha = x_1 \quad \text{and} \quad \beta = \frac{1}{\bar{x} - x_1}.$$

Note:

1. Whenever the given probability function involves a constant and the range of the variable is dependent on the parameters to be estimated, then we have to determine the constant by taking the total probability as unity and then proceed with the estimation part.

2. From Examples (8) and (9), it is understood that whenever the range of the variable involves parameters to be estimated, the likelihood equations fail to give valid estimates and M.L.E. are obtained by following some other methods.

3.11 THE ESTIMATION OF MEANS

In section 3.2 we dealt with point estimation. It does not reveal on how much information the estimate is based nor does it tell anything about the size of the error. Hence we have to supplement a point estimate $\hat{\theta}$ of θ with the size of the sample and the value of Var $[\hat{\theta}]$ or with some other information about the sampling distribution of $\hat{\theta}$.

An interval estimate of θ is an interval of the form $\hat{\theta}_1 < \theta < \hat{\theta}_2$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the values of appropriate random variables $\hat{\theta}_1$ and $\hat{\theta}_2$.

$P\,[\hat{\theta}_1 < \theta < \hat{\theta}_2] = 1 - \alpha$, for probability $1 - \alpha$. For a specific value of $1 - \alpha$, it is referred to $\hat{\theta}_1 < \theta < \hat{\theta}_2$ as a $(1 - \alpha)\, 100\%$ confidence interval for θ. Here $(1 - \alpha)$ is called the degree of confidence and the ends of the interval $\hat{\theta}_1$ and $\hat{\theta}_2$ are called the lower and upper confidence limits. It is noted that, like point estimates, interval estimates of a given parameter are not unique. The methods of interval estimation are judged by their various statistical properties.

Suppose that the mean of a random sample is to be used to estimate the mean of a normal population with known variance σ^2. The sampling distribution of $\overline{X}$ for random samples of size n from a normal population with mean μ and variance σ^2 is a normal distribution with $\mu \bar{x} = \mu$ and $\sigma^2 = \frac{\sigma^2}{n}$.

Then we know that from section 3.4 maximum error of estimate

$$P\left[-t_{\alpha/2} \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha/2}\right] = 1-\alpha\,.$$

$$\Rightarrow \qquad P\left[\frac{|\bar{x}-\mu|}{\left(\frac{\sigma}{\sqrt{n}}\right)} \leq Z_{\alpha/2}\right] = 1-\alpha$$

$$\text{(or)} \qquad P\left[|\bar{x}-\mu| \leq Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha\,.$$

Theorem 8: If $\bar{X}$ is the mean of a random sample of size n from a normal population with mean μ and variance σ^2, its sampling distribution is a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Theorem 9: If $\bar{X}$, the mean of a random sample of size "n" from a normal population, with known variance σ^2, is to be used as an estimator of those mean of the population, the probability is $(1-\alpha)$ that the error will be less than

$$Z_{\alpha/2}\cdot\left(\frac{\sigma}{\sqrt{n}}\right).$$

Theorem 10: If $\bar{x}$ is the value of the mean of a random sample of size n from a normal population with known variance σ^2, then

$$\bar{x}-Z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+t_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}},$$

is a $(1-\alpha)$ 100% confidence interval for the mean of the population.

Theorem 11: If $\bar{x}$ and s are the values of the mean and S.D of a random sample of size n from a normal population, then

$$\bar{x}-t_{\alpha/2,\,n-1}\cdot\frac{s}{\sqrt{n}} < \mu < \bar{x}+t_{\alpha/2,\,n-1}\cdot\frac{s}{\sqrt{n}}$$

is a $(1-\alpha)$ 100% confidence interval for the mean of the population.

WORKED EXAMPLES

Example: 1

A team of efficiency experts intends to use the mean of a random sample of size n = 150 to estimate the average mechanical aptitude to assembly-line workers in a large industry. It, based on experience, the efficiency experts can assume that σ = 6.2 for such data, what can they assert with probability 0.99 about the maximum error of their estimate?

Solution:

Given that $n = 150, \sigma = 6.2$ and $z_{0.005} = 2.575$, substitute these values in the equation,

$$Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \text{ we have } 2.575 \times \frac{6.2}{\sqrt{150}} = 1.30.$$

Hence the efficiency experts can assert with probability 0.99 that their error will be less than 1.30.

Example: 2

If a random sample of size n = 20 from a normal population with the variance σ² = 225 has the mean x̄ = 64.3, construct a 95% confidence interval for the population mean μ.

Solution:

It is given that $n = 20, \bar{x} = 64.3, \sigma^2 = 225$ and $Z_{0.025} = 1.96$.

We know that, if $\bar{x}$ and σ are known, then the confidence interval formula is

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$64.3 - 1.96 \cdot \frac{15}{\sqrt{20}} < \mu < 64.3 + 1.96 \cdot \frac{15}{\sqrt{20}}$$

$$\Rightarrow 57.7 < \mu < 70.9$$

Example: 3

A paint manufacturer wants to determine the average drying time of a new interior wall paint. If for 12 test areas of equal size he obtained a mean drying time of 66.3 minutes and a standard deviation of 8.4 minutes, construct a 95% confidence interval for the true mean μ.

Solution:

It is given that

$$\bar{x} = 66.3, n = 12, s = 8.4 \text{ and } t_{0.025,\, 11} = 2.201.$$

If $\bar{x}$ and s are known then, the 95% confidence interval for μ is given by

$$\bar{x} - t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$$

$$66.3 - 2.201 \cdot \frac{8.4}{\sqrt{12}} < \mu < 66.3 + 2.201 \cdot \frac{8.4}{\sqrt{12}}$$

$\Rightarrow \; 61.0 < \mu < 71.6$

That means we can assert with 95% confidence that the interval from 61.0 minutes to 71.6 minutes contains the true average drying time of the paint.

Example: 4

A district official intends to use the mean $\bar{x} = 61.8$ of a random sample of 150 sixth graders from a very large school district to estimate the mean score which all the sixth graders in the district would get if they took a certain arithmetic achievement test. If, based on experience, the official knows that $\sigma = 9.4$ for such data, what can she assert with probability 0.99 about the maximum error?

Solution:

It is given that $\bar{x} = 61.8, \sigma = 9.4, n = 150$ and $Z_{0.005} = 2.575$.

We know that if $\bar{x}$ and σ is given then the confidence interval formula is

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad 61.8 - 2.575 \cdot \frac{9.4}{\sqrt{150}} < \mu < 61.8 + 2.575 \cdot \frac{9.4}{\sqrt{150}}$$

$$\Rightarrow \quad 61.8 - 1.9764 < \mu < 61.8 + 1.9764$$

$$\therefore \quad 59.8236 < \mu < 63.7764.$$

Thus she can assert with 99% confidence that the interval from).8236 to 63.7764.

:xample: 5

A medical research worker intends to use the mean of a random imple of size $n = 120$ with $\bar{x} = 141.8$ mm of mercury to estimate the iean blood pressure of women in their fifties. If, based on experience, e knows that $\sigma = 10.5$ mm of mercury. Construct a 99% confidence iterval for the mean blood pressure of women in their fifties.

Solution:

It is given that $n = 120, \bar{x} = 141.8$ and $\sigma = 10.5$, $Z_{0.005} = 2.575$.

We know that if $\bar{x}$ and σ are given then the confidence interval ormula is given by

$$\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$141.8 - 2.575 \cdot \frac{10.5}{\sqrt{120}} < \mu < 141.8 + 2.575 \frac{10.5}{\sqrt{120}}$$

$$141.8 - 2.46817 < \mu < 141.8 + 2.46817$$

$$139.3318 < \mu < 144.2682$$

Example: 6

A major truck shop has kept extensive records on various transactions with its customers. If a random sample of 18 of these records show average sales of 63.84 gallons of diesel fuel with a S.D of 2.75 gallons, construct 99% confidence interval for the mean of the population sampled.

Solution:

It is given that $n = 18, \bar{x} = 63.84, \sigma = 2.75$ and $Z_{0.005} = 2.575$.

If $\bar{x}$ and σ are known, then the confidence interval is

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow 63.84 - 2.515 \cdot \frac{2.75}{\sqrt{18}} < \mu < 63.84 + 2.575 \frac{2.75}{\sqrt{18}}$$

$$63.84 - 1.6691 < \mu < 63.84 + 1.6691$$

$$62.1709 < \mu < 65.5091$$

3.12 THE ESTIMATION OF DIFFERENCES BETWEEN MEANS

From normal populations, we can find for independent random samples

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

has the standard normal distribution. If we substitute the above equation in

$$P[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1 - \alpha,$$

the pivotal method gives the confidence interval formula for $\mu_1 - \mu_2$.

If $\bar{x}_1$ and $\bar{x}_2$ are the values of the means of independent random samples of sizes n_1 and n_2 from normal populations with known variances σ_1^2 and σ_2^2, then

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

is a $(1 - \alpha)$ 100% confidence interval for the difference between the two population means.

This formula can also be used for independent random sample from non-normal populations with known variances when n_1 and n_2 are large samples (ie. $n_1 \geq 30$ and $n_2 \geq 30$).

To construct a $(1-\alpha)$ 100% confidence interval for the difference between two means when $n_1 \geq 30, n_2 \geq 30$, but σ_1 and σ_2 are unknown, we simply substitute s_1 and s_2 for σ_1 and σ_2 and then we have to proceed. When σ_1 and σ_2 are unknown and either or both of the samples are small, the procedure for estimating the difference between the means of two normal populations is not straightforward unless it can be assumed as $\sigma_1 = \sigma_2$.

If $\sigma_1 = \sigma_2 = \sigma$ then

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is a random variable having the standard normal distribution and σ^2 can be estimated by pooling the squared deviations from the means of the two samples. Then the pooled estimator is defined by

$$S_p^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2. Now the independent random variables,

$$\frac{(n_1 - 1)\, S_1^2}{\sigma^2} \text{ and } \frac{(n_2 - 1)\, S_2^2}{\sigma^2}$$

have Chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, and their sum is given by

$$Y = \frac{(n_1 - 1)\, S_1^2}{\sigma^2} + \frac{(n_2 - 1)\, S_2^2}{\sigma^2} = \frac{(n_1 + n_2 - 2)}{\sigma^2}\, S_p^2$$

has a Chi-square distribution with $n_1 + n_2 - 2$ degrees of freedom.

If the random variables Z and Y are independent then

$$T = \frac{Z}{\sqrt{\frac{Y}{n_1 + n_2 - 2}}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has t - distribution with $n_1 + n_2 - 2$ degrees of freedom. Substituting this expression for T into

$$P[-t_{\alpha/2,\, n-1} < T < t_{\alpha/2,\, n-1}] = 1 - \alpha .$$

We have the following $(1-\alpha)$ 100% confidence interval for $\mu_1 - \mu_2$.

If $\bar{x}_1, \bar{x}_2, s_1$ and s_2 are the values of the means and the standard deviations of independent random samples of sizes n_1 and n_2 from normal populations with equal variances, then

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,\, n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,\, n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

is a $(1-\alpha)$ 100% confidence interval for the difference between the two populations means. Since this confidence interval formula is used when n_1 and/or n_2, are small less than 30, we have to use the small sample confidence interval for $\mu_1 - \mu_2$.

Example: 7

Construct a 94% confidence interval for the difference between the mean life-time of two kinds of light bulbs, given that a random sample of 40 light bulbs of the first kind lasted on the average 418 hours of continuous use and 50 light bulbs of the second kind lasted on the average 402 hours of continuous use. The population S.D are known to be $\sigma_1 = 26$ and $\sigma_2 = 22$.

Solution:

For $\alpha = 0.06$ the table value of $Z_{0.03} = 1.88$. It is given that $n_1 = 40$, $n_2 = 50$, $\bar{x}_1 = 418$, $\bar{x}_2 = 402$, $\sigma_1 = 26$ and $\sigma_2 = 22$.

We know that if $\bar{x}_1, \bar{x}_2, \sigma_1^2$ and σ_2^2 are known then the corresponding confidence interval is

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_0 <$$

$$(\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$\Rightarrow \quad (418-402) - 1.88\sqrt{\frac{(26)^2}{40}+\frac{(22)^2}{50}} < \mu_1 - \mu_2$$

$$< (418-402) + 1.88\sqrt{\frac{(26)^2}{40}+\frac{(22)^2}{50}}$$

$$(16) - 1.88\sqrt{16.9+9.68} < \mu_1 - \mu_2 < (16) + 1.88\sqrt{16.9+9.68}$$

$$(16) - 1.88\,(5.1555) < \mu_1 - \mu_2 < (16) + 1.88\,(5.1555)$$

$$16 - 9.69234 < \mu_1 - \mu_2 < 16 + 9.69234$$

$$6.30766 < \mu_1 - \mu_2 < 25.69234$$

Hence, we are 94% confident that the interval from 6.30766 to 25.69234 hours contains the actual difference between lifetimes of the two kinds of light bulbs. The fact that both confidence limits are positive suggests that on the average the first kind of light bulb is superior to the second kind.

Example: 8

Independent random samples of sizes $n_1 = 16, n_2 = 25$ from normal populations with $\sigma_1 = 4.8$ and $\sigma_2 = 3.5$ have the means $\bar{x}_1 = 18.2$ and $\bar{x}_2 = 23.4$. Find a 95% confidence interval for $\mu_1 - \mu_2$.

Solution:

It is given that $n_1 = 16, n_2 = 25$,

$$\bar{x}_1 = 18.2, \bar{x}_2 = 23.4, \sigma_1 = 4.8, \sigma_2 = 3.5 \text{ and } Z_{0.005} = 2.575.$$

We know that

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\Rightarrow (18.2 - 23.4) - 2.575\sqrt{\frac{(4.8)^2}{16} + \frac{(3.5)^2}{25}} < \mu_1 - \mu_2$$

$$< (18.2 - 23.4) + 2.575\sqrt{\frac{(4.8)^2}{16} + \frac{(3.5)^2}{25}}$$

$$(-5.2) - 2.575\sqrt{1.44 + 0.49} < \mu_1 - \mu_2 < (-5.2) + 2.575\sqrt{1.44 + 0.49}$$

$$(-5.2) - 2.575\,(1.3892) < \mu_2 - \mu_1 < (-5.2) + 2.575\,(1.3892)$$

$$-5.2 - 3.57719 < \mu_1 - \mu_2 < (-5.2) + 3.57719$$

$$-8.77719 < \mu_1 - \mu_2 < -1.62281$$

Example: 9

A study has been made to compare the nicotine contents of two brands of cigarettes. Ten cigarettes of Brand A has an average nicotine content at 3.1 mg with a S.D of 0.5 mg, while eight cigarettes of brand B had an average nicotine content of 2.7 m.g with a S.D of 0.7 m.g. Assuming that the two sets of data, one independent random samples from normal populations with equal variances, construct a 95% confidence interval for the difference between the mean nicotine contents of the two brands of cigarettes.

✍ **Solution:**

It is given that $n_1 = 10$, $n_2 = 8$, $s_1 = 0.5$ and $s_2 = 0.7$. We know that the formula to find S_p is

$$S_p^2 = \frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}$$

$$\Rightarrow \quad S_p = \sqrt{\frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-2}} = \sqrt{\frac{9\,(0.5)^2 + 7\,(0.49)}{10+8-2}}$$

$$= \sqrt{\frac{9\,(0.25) + 7\,(0.49)}{16}} = \sqrt{\frac{2.25+3.43}{16}} = \sqrt{\frac{16}{5.68}}$$

$$\Rightarrow \quad S_p = 0.596$$

$$\therefore\ n_1 = 10,\ n_2 = 8,\ s_1 = 0.5,\ s_2 = 0.7,\ S_p = 0.596,$$

$$\bar{x}_1 = 3.1,\ \bar{x}_2 = 2.7 \text{ and } t_{0.025,\,16} = 2.120.$$

Then we know that, if $\bar{x}_1$, $\bar{x}_2$ and S_p are known then the corresponding confidence interval is

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,\,n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 <$$

$$(\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,\,n_1+n_2-2} \cdot S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$(3.1-2.7) - 2.120\,(0.596)\,\sqrt{\frac{1}{10}+\frac{1}{8}} < \mu_1 - \mu_2 <$$

$$(3.1-2.7) + 2.120\,(0.596\sqrt{\frac{1}{10}+\frac{1}{8}}$$

$$\Rightarrow \quad (0.4) - (1.26352)\,(0.4743) < \mu_1 - \mu_2$$

$$< (0.4) + (1.26352)\,(0.4743)$$

$$\Rightarrow \quad -0.1992 < \mu_1 - \mu_2 < 0.9992$$

Thus the 95% confidence limits are -0.1992 and 0.9992 m.g. But here we observe that since this include $\mu_1 - \mu_2 = 0$, we cannot conclude that there is a real difference between the average nicotine contents of the two brands of cigarettes.

Example: 10

Twelve randomly selected mature citrus trees of one variety hav mean height of 13.8 feet with a S.D of 1.2 feet, and fifteen rando selected mature citrus trees of another variety have a mean height 12.9 feet with a S.D of 1.5 feet. Assuming that the random samples w selected from normal populations with equal variances, construct a 9 confidence interval for the difference between the true average heig of the two kinds of citrus trees.

Solution:

It is given that $n_1 = 12,\ n_2 = 15,$

$$\bar{x}_1 = 13.8,\ \bar{x}_2 = 12.9,\ s_1 = 1.2 \text{ and } s_2 = 1.5$$

The tabulated value is $t_{0.025,\,25} = 2.060.$

We know that

$$S_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}}$$

$$S_p = \sqrt{\frac{11\,(1.2)^2 + 14\,(1.5)^2}{12 + 15 - 2}} = \sqrt{\frac{15.84 + 31.5}{25}}$$

$$S_p = 1.3761$$

If $\bar{x}_1, \bar{x}_2$ and S_p are known, then the corresponding confidence inter

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,\, n_1 + n_2 - 2} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,\, n_1 + n_2 - 2} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(13.8 - 12.9) - 2.060\,(1.3761) \sqrt{\frac{1}{12} + \frac{1}{15}} < \mu_1 - \mu_2$$

$$< (13.8 - 12.9) + 2.060\,(1.3761) \sqrt{\frac{1}{12} + \frac{1}{15}}$$

$$(0.9) - 2.060\ (1.3761)\ (0.3873) < \mu_1 - \mu_2 < (0.9) + 2.060\ (1.3761)\ (0.3873)$$

$$0.9 - 1.0979 < \mu_1 - \mu_2 < 0.9 + 1.0979$$

$$-0.1979 < \mu_1 - \mu_2 < 1.9979$$

3.13 THE ESTIMATION OF PROPORTIONS

There are many problems in which we must estimate proportions, probabilities, percentages, rates such as the proportions of detectives in a large shipment of transistors, the probability that a car stopped at a road block will have faulty lights, the mortality rate of a disease. In these situations, we are sampling a binomial population and hence that our problem is to estimate the binomial parameter θ. For large n, the binomial distribution can be approximated with a normal distribution, that

$$Z = \sqrt{\frac{X - n\,\theta}{n\,\theta\,(1-\theta)}}$$

can be treated as a random variable having approximately the standard normal distribution. Substituting this expression for Z into

$$P\,[-Z_{\alpha/2} < Z < Z_{\alpha/2}] = 1 - \alpha.$$

We get

$$P\left[-Z_{\alpha/2} < \frac{X - n\,\theta}{\sqrt{n\,\theta\,(1-\theta)}} < Z_{\alpha/2}\right] = 1 - \alpha$$

and the two inequalities

$$-Z_{\alpha/2} < \frac{x - n\,\theta}{\sqrt{n\,\theta\,(1-\theta)}} \quad \text{and} \quad \frac{x - n\,\theta}{\sqrt{n\,\theta\,(1-\theta)}} < Z_{\alpha/2},$$

whose solution will give $(1-\alpha)$ 100% confidence limits for θ. If X is a binomial random variable with the parameters n and θ, n is large and $\hat{\theta} = \frac{x}{n}$ then

$$\hat{\theta} - Z_{\alpha/2}\sqrt{\frac{\hat{\theta}\,(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + Z_{\alpha/2}\cdot\sqrt{\frac{\hat{\theta}\,(1-\hat{\theta})}{n}}$$ is an approximate

$(1-\alpha)$ 100% confidence interval for θ.

Note:

If $\hat{\theta} = \frac{x}{n}$ is used as an estimate of θ, we can assert with $(1 - \alpha)$ 100% confidence that the error is less than

$$Z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}.$$

3.14 THE ESTIMATION OF DIFFERENCES BETWEEN PROPORTIONS

There are many problems in which we must estimate the difference between the binomial parameters θ_1 and θ_2 on the basis of independent random samples of sizes n_1 and n_2 from two binomial populations.

If X_1 is a binomial random variable with parameters n_1 and θ_1, X_2 is a binomial random variable with the parameters n_2 and θ_2, when n_1 and n_2 are large, and $\hat{\theta}_1 = \frac{x_1}{n_1}$ and $\hat{\theta}_2 = \frac{x_2}{n_2}$ then,

$$(\hat{\theta}_1 - \hat{\theta}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}} < \theta_1 - \theta_2$$

$$< (\hat{\theta}_1 - \hat{\theta}_2) + Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}$$

is an approximate $(1-\alpha)$ 100% confidence interval for $\theta_1 - \theta_2$.

Example: 11

In a random sample 136 of 400 persons given a flu vaccine experienced some discomfort. Construct a 95% confidence interval for the true proportion of person who will experience some discomfort from the vaccine.

Solution:

It is given that $n = 400$, $\hat{\theta} = \frac{x}{n} = \frac{136}{400} = 0.34$ and $Z_{0.025} = 1.96$.

We know that

$$\hat{\theta} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

$$(0.34) - 1.96\sqrt{\frac{(0.34)(0.66)}{400}} < \theta < (0.34) + 1.96\sqrt{\frac{(0.34)(0.66)}{400}}$$

$$(0.34) - 1.96\,(0.0237) < \theta < (0.34) + 1.96\,(0.0237)$$

$$0.2935 < \theta < 0.3865$$

Example: 12

A study is made to determine the proportion of voters in a sizeable community who favour the construction of a nuclear power plant. If 140 of 400 voters selected at random favour the project and we use $\hat{\theta} = \frac{140}{400} = 0.35$ as an estimate of the actual proportion of all voters in the community who favour the project, what can we say with 99% confidence about the maximum error?

Solution:

It is given that $n = 400$, $\hat{\theta} = 0.35$ and $Z_{0.005} = 2.575$.

We know that if $\hat{\theta} = \frac{x}{n}$ is used as an estimate of θ, with $(1-\alpha)$ 100% confidence that the error is less than

$$Z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}.$$

$$\therefore\ Z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = 2.575\sqrt{\frac{(0.35)(0.65)}{400}} = 0.061.$$

Hence, if we use $\hat{\theta} = 0.35$ as an estimate of the actual proportion of voters in the community who favour the project, we can assert with 99% confidence that the error is less than 0.061.

Example: 13

A sample survey of a supermarket showed that 204 of 300 shoppers regularly use cents-off coupons. Construct a 95% confidence interval for the corresponding true proportion.

Solution:

Give that $n = 300, \hat{\theta} = \frac{x}{n} = \frac{204}{300} = 0.68$ and $Z_{0.025} = 1.96$.

We know that

$\hat{\theta} - Z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + Z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$ is an approximate $(1-\alpha)$ 100% confidence interval for θ.

$$(0.68) - 1.96\sqrt{\frac{(0.68)(0.32)}{300}} < \theta < (0.68) + 1.96\sqrt{\frac{(0.68)(0.32)}{300}}$$

$$(0.68) - 1.96(0.0269) < \theta < (0.68) + 1.96(0.0269)$$

$$0.6272 < \theta < 0.7327$$

Example: 14

A sample survey at a supermarket showed that 204 of 300 shoppers regularly use cents-off coupons. What can we say with 99% confidence about the maximum error, if we use the observed sample proportion as an estimate of the proportion of all shoppers in the population sampled who use cents-off coupons?

Solution:

Given that $n = 300, \hat{\theta} = \frac{x}{n} = \frac{204}{300} = 0.68$, and $Z_{0.005} = 2.575$.

We know that $Z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

$$\Rightarrow \quad 2.575\sqrt{\frac{(0.68)\,(0.32)}{300}} = 2.575\,(0.0269) = 0.069$$

$\therefore$ 99% confidence about the maximum error is 0.069.

Example: 15

If 132 of 200 male voters and 90 of 150 female voters favour a certain candidate running for governor of India, find a 99% confidence interval for the difference between the actual proportions of male and female voters who favour the candidate.

Solution:

Given that $\hat{\theta}_1 = \frac{132}{200} = 0.66, n_1 = 200, n_2 = 150, x_1 = 132, x_2 = 90,$

$$\hat{\theta}_2 = \frac{90}{150} = 0.60 \text{ and } Z_{0.005} = 2.575.$$

We know that if $\hat{\theta}_1 = \frac{x_1}{n_1}$ and $\hat{\theta}_2 = \frac{x_2}{n_2}$ are given then

$$(\hat{\theta}_1 - \hat{\theta}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1\,(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2\,(1-\hat{\theta}_2)}{n_2}} < \theta_1 - \theta_2$$

$$< (\hat{\theta}_1 - \hat{\theta}_1) + Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1\,(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2\,(1-\hat{\theta}_2)}{n_2}}$$

$$\Rightarrow (0.66 - 0.60) - 2.575\sqrt{\frac{(0.66)\,(0.34)}{200} + \frac{(0.60)\,(0.40)}{150}} < \theta_1 - \theta_2$$

$$< (0.66 - 0.60) + 2.575\sqrt{\frac{(0.66)\,(0.34)}{200} + \frac{(0.60)\,(0.40)}{150}}$$

$$\Rightarrow (0.06) - 2.575\,\sqrt{0.0011 + 0.0016} < \theta_1 - \theta_2$$

$$< (0.06) + 2.575\,\sqrt{0.0011 + 0.0016}$$

$$\Rightarrow (0.06) - 2.575\,(0.05196) < \theta_1 - \theta_2 < (0.06) + 2.575\,(0.05196)$$

$$-0.0737 < \theta_1 - \theta_2 < 0.1937\,.$$

We are 99% confident that the interval from -0.0737 to 0.1937 contains the difference between the actual proportions of male and female voters into favour the candidate. This includes the possibility of a zero difference between the two proportions.

Example: 16

In a random sample of visitors to a famous tourist attractions 84 of 250 men and 156 of 250 women bought souvenirs. Construct a 95% confidence interval for the difference between the true proportions of men and women who buy souvenirs at this tourist attraction.

Solution:

Given that $n_1 = 250$, $n_2 = 250$, $x_1 = 84$,

$x_2 = 156$, $z_{0.025} = 1.96$, $\hat{\theta}_1 = \dfrac{84}{250} = 0.336$ and $\hat{\theta}_2 = \dfrac{156}{250} = 0.624$.

If $\hat{\theta}_1 = \dfrac{x_1}{n_1}$ and $\hat{\theta}_2 = \dfrac{x_2}{n_2}$ are known, we know that

$$(\hat{\theta}_1 - \hat{\theta}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}} < \theta_1 - \theta_2$$

$$< (\hat{\theta}_1 - \hat{\theta}_2) + Z_{\alpha/2}\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}$$

$$\Rightarrow (0.336 - 0.624) - 1.96\sqrt{\frac{(0.336)\,(0.664)}{250} + \frac{(0.624)\,(0.376)}{250}}$$

$$< \theta_1 - \theta_2 < (0.336 - 0.624) + 1.96\sqrt{\frac{(0.336)\,(0.664)}{250} + \frac{(0.624)\,(0.376)}{250}}$$

$$\Rightarrow (-0.288) - 1.96\sqrt{0.00089 + 0.00093} < \theta_1 - \theta_2$$
$$< (-0.288) + 1.96\sqrt{0.00089 + 0.00093}$$
$$\Rightarrow (-0.288) - 1.96\,(0.04266) < \theta_1 - \theta_2 < (-0.288) + 1.96\,(0.04266)$$
$$\Rightarrow -0.3716 < \theta_1 - \theta_2 < -0.2043.$$

Hence we are 95% confident that the interval from -0.3716 to -0.2043 contains the difference between the true proportions of men and women who buy souvenirs at the tourist attraction.

3.15 THE ESTIMATION OF VARIANCES

If $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the standard deviation σ then

(i) $\overline{X}$ and S^2 are independent.

(ii) the random variable $\dfrac{(n-1)\,S^2}{\sigma^2}$ has a Chi-square distribution with $(n-1)$ degrees of freedom.

Based on the above concept, given a random sample of size n from a normal population, we can obtain a $(1-\alpha)$ 100% confidence interval for σ^2, according to which

$$\frac{(n-1)\,S^2}{\sigma^2}$$

is a random variable having a Chi-square distribution with $(n-1)$ degrees of freedom. Then

$$P\left[\chi^2_{1-\alpha/2,\,n-1} < \frac{(n-1)\,S^2}{\sigma^2} < \chi^2_{\alpha/2,\,n-1}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}\right] = 1-\alpha.$$

Thus if S^2 is the value of the variance of a random sample of size n from a normal population, then

$$\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)^2\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

is a $(1-\alpha)$ 100% confidence interval for σ^2. Corresponding $(1-\alpha)$ 100% confidence limits for σ can be obtained by taking the square roots of the confidence limits for σ^2.

3.16 THE ESTIMATION OF THE RATIO OF TWO VARIABLES

If S_1^2 and S_2^2 are the variances of independent random samples of sizes n_1 and n_2 from normal population, then

$$F = \frac{\sigma_2^2\,S_1^2}{\sigma_1^2\,S_2^2}$$

is a random variable having an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

$$P\left[f_{1-\alpha/2,\,n_1-1,\,n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2,\,n_2-1,\,n_1-1}\right] = 1-\alpha.$$

Since $f_{1-\alpha/2,\,n_1-1,\,n_2-1} = \dfrac{1}{f_{\alpha/2,\,n_1-1,\,n_2-1}}$, then

$$\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\alpha/2,\,n_1-1,\,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\,f_{\alpha/2,\,n_2-1,\,n_1-1}\,.$$

is a $(1-\alpha)$ 100% confidence interval for $\dfrac{\sigma_1^2}{\sigma_2^2}$.

Corresponding $(1-\alpha)$ 100% confidence limits for $\dfrac{\sigma_1}{\sigma_2}$ can be obtained by taking the square roots of the confidence limits for $\dfrac{\sigma_1^2}{\sigma_2^2}$.

Example: 17

In 16 test runs the gasoline consumption of an experiment engine has a standard deviation of 2.2 gallons. Construct a 99% confidence interval for σ^2, which measures the true variability of the gasoline consumption of the engine.

Solution:

Let us assume that the given data as a random sample from a normal population. It is given that $n = 16$, $S = 2.2$. Since σ and n are given, then the corresponding confidence interval is

$$\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

Here $\chi^2_{\alpha/2,\,n-1} = \chi^2_{0.005,\,15} = 32.801$ and $\chi^2_{1-\alpha/2,\,n-1} = \chi^2_{0.995,\,15} = 4.601$.

$$\therefore \quad \frac{(16-1)\,(2.2)^2}{32.801} < \sigma^2 < \frac{(16-1)\,(2.2)^2}{4.601}$$

$$\frac{15\,(4.84)}{32.801} < \sigma^2 < \frac{15\,(4.84)}{4.601}$$

$$2.2133 < \sigma^2 < 15.7792$$

To get a corresponding 99% confidence interval for σ, we take square roots and get

$$1.49 < \sigma < 3.97$$

Example: 18

The length of the skulls of 10 fossil skeletons of an extinct species of birds has a mean of 5.68 cm and a S.D of 0.29 cm. Assuming that such measurements are normally distributed, construct a 95% confidence interval for the variance of skull length of the given species of birds.

Solution

Give that $n = 10$, $s = 0.29$, $\bar{x} = 5.68$.

We know that $\frac{(n-1)\,S^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2,\,n-1}}$

From the table $\chi^2_{\alpha/2,\,n-1} = \chi^2_{0.025,\,9} = 19.023$ and

$$\chi^2_{1-\alpha/2,\,n-1} = \chi^2_{0.975,\,9} = 2.7.$$

$$\Rightarrow \quad \frac{9\,(0.29)^2}{19.023} < \sigma^2 < \frac{9\,(0.29)^2}{2.7}$$

$$\Rightarrow \quad \frac{0.7569}{19.023} < \sigma^2 < \frac{0.7569}{2.7}$$

$$\Rightarrow \quad 0.0398 < \sigma^2 < 0.2803$$

Example: 19

A study has been made to compare the nicotine contents of two brands of cigarettes. Ten cigarettes of brand A had an average nicotine content of 3.1 mg with S.D of 0.5 mg, while eight cigarettes of brand B has an average nicotine content of 2.7 mg with a S.D of 0.7 mg. Assuming that the two sets of data are independent random samples, from normal population with equal variances, construct a 98% confidence interval for σ_1^2/σ_2^2.

Solution:

Give that $n_1 = 10, n_2 = 8, S_1 = 0.5, S_2 = 0.7$

$$f_{\alpha/2,\,n_1-1,\,n_2-1} = f_{0.01,\,9,\,7} = 6.72 \text{ and}$$

$$f_{\alpha/2,\,n_2-1,\,n_1-1} = f_{0.01,\,7,\,9} = 5.61.$$

If S_1^2 and S_2^2 are the values of the variances of independent random samples of sizes n_1 and n_2, then

We know that $\frac{S_1^2}{S_2^2}\,\frac{1}{f_{\alpha/2,\,n_1-1,\,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2,\,n_2-1,\,n_1-1}$

$$\Rightarrow \frac{(0.5)^2}{(0.7)^2}\cdot\frac{1}{6.72} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{(0.5)^2}{(0.7)^2}\,5.61$$

$$\Rightarrow 0.076 < \frac{\sigma_1^2}{\sigma_2^2} < 2.862$$

UNIT-4 NON-PARAMETRIC TESTS

1. Sign Test

Definition:-

→ The sign Test is a statistical method to test for consistent difference between pairs of observations and not on their numerical magnitudes.

→ It is an easiest non-parametric Test.

formula:-

$$K = \frac{n-1}{2} - (0.98\sqrt{n})$$

where, n = no. of attributes with sign (Zero Neglected)

K = Calculated value.

Hypothesis resultant tests:

If $S > K$; Hypothesis is Accepted

If $S < K$; Hypothesis is Rejected.

where S – Table value calculated by Total numbers of negatives

k – Calculated value (using formula)

Problem 1 :

Use sign Test to see if there is a difference b/w No. of days until collection of an account receivable before & after a new collection policy.

Before	30	28	34	35	40	42	33	38	34
After	32	29	33	32	37	43	40	41	37

Before	45	28	27	25	41	36
After	44	27	33	30	38	36

Soln:-

Before	After	Sign = Before − After.
30	32	−
28	29	−
34	33	+
35	32	+
40	37	+
42	43	−
33	40	−
38	41	−
34	37	−
45	44	+
28	27	+
27	33	

Before	After	Sign = Before − After
25	30	−
41	38	+
36	36	0

$S = 8$ (No of −ve)

Table Value.

$S =$ No. of Negatives $= 8$

$$\boxed{\therefore S = 8}$$

Calculated Value:

Given $n = 14$ (without zero)

$$K = \frac{n-1}{2} - (0.98\sqrt{n})$$

$$= \frac{14-1}{2} - (0.098\sqrt{14})$$

$$\boxed{K = 2.83}$$

Conclusion:

Since $S > K$, Hypothesis is accepted.

Table value ($S=8$) > Cal. value ($K = 2.83$)

$\therefore$ There is no significant difference before & after a new collection of policy in accounts receivable.

Problem – 2:

The following data constitute a random sample of 15 measurements of the octane rating of a certain kind of gasoline:

99.0	102.3	99.8	100.5	99.7	96.2	99.1	102.5
103.3	97.4	100.4	98.9	98.3	98.0	101.6	

Test the null hypothesis $\tilde{\mu} = 98.0$ against the alternative hypothesis $\tilde{\mu} > 98.0$ at the 0.01 level of significance.

Soln:

Step 1: Null hypothesis : $\tilde{\mu} = 98.0$ $\left(p = \frac{1}{2}\right)$

Alternative hypothesis : $\tilde{\mu} > 98.0$ $\left(p > \frac{1}{2}\right)$

Step 2:

Level of significance : $\alpha = 0.01$

Step 3:

Criterion : The criterion may be based on the number of plus signs or the number of minus signs. Using the number of plus signs, denoted by x, reject the null hypothesis if the probability of getting x or more plus signs is less than or equal to 0.01.

Step 4:

Calculations : Replacing each value greater than 98.0 with a plus sign and each value less than 98.0 with a minus sign, the 14 samples values yeild.

+ + + + + − + +

+ − + + + + +

Thus $x = 12$ and a table of binomial distribution shows that for $n = 14$ and $p = 0.50$ the probability of $X \geq 12$ is $1 - (x < 12) = 1 - 0.9935 = \boxed{0.0065}$

Step 5:

Since 0.0065 is less than 0.01, the null hypothesis must be rejected; we conclude that the median octane rating of the given kind of gasoline exceeds 98.0.

2. One Sample Run Test :-

* One sample run test is used to identify a non-random pattern in a sequence of elements.

Formula:

$$Z = \frac{r - \mu_r}{\sigma_r}$$

where r = No of groups available

$$\mu_r = \frac{2 n_1 n_2}{n_1 + n_2} + 1 \quad ; \mu_r - \text{Mean of } r$$

where n_1 - no of elements in 1^{st} group

n_2 - no of elements in 2^{nd} group

$\sigma_r \Rightarrow$ Standard error of "r"

$$\sigma_r = \sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

Significance value:

$1\% = 2.58$; $5\% = 1.96$

If $Z <$ table value, Hypothesis is accepted

$Z >$ table value, Hypothesis is Rejected.

Problem 1:

The following is an arrangement of 25 Men (M) and 15 Women (W), lined up to purchase tickets for a premier picture show:-

M WW MMM W MM W M W M WWW MMM W MM WWW MMM MMM WWW MMMMMM

Test for randomness at the 5% significance.

Soln: Null Hypothesis : Arrangement of samples in Random

Alternative Hypothesis: There is a frequency alternating patterns

Level of significance: 0.05

Given:

n_1 = No. of Mens (M) = 25 ; $n_1 = 25$

n_2 = No. of Women (W) = 15 ; $n_2 = 15$

r = No. of groups = 17 ; $r = 17$

formula:

$$Z = \frac{r - \mu_r}{\sigma_r} \quad ; \quad Z \rightarrow \text{Calculated value.}$$

$$\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 \Rightarrow \mu_r = \frac{2(25)(15)}{25+15} + 1 \Rightarrow \mu_r = \frac{750}{40} + 1$$

$$\boxed{\therefore \mu_r = 19.75}$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}} \Rightarrow \sigma_r = \sqrt{\frac{2(25)(15)(2(25)(15) - 25 - 15)}{(25+15)^2 (25+15-1)}}$$

$$\sigma_r = \sqrt{\frac{750(710)}{40^2(39)}} \Rightarrow \boxed{\sigma_r = 2.92}$$

$$Z = \frac{r - \mu_r}{\sigma_r} \Rightarrow \frac{17 - 19.75}{2.92} \Rightarrow Z = -0.94 \text{ (sign ignored)}$$

$$\boxed{\therefore Z = 0.94}$$

Table value:

Level of significance $= \alpha = 0.05 = 1.96$.

Conclusion:

Calculated value

$\therefore Z\ (0.94) <$ Table Value (1.96)

Hence the Hypothesis is accepted. Hence, there is no real evidence to suggest that arrangement is not random.

Problem: 2

An engine is concerned about the possibility that too many chances are being made in this setting of a Automatic given the following Mean diameters (inches) of 40 successive shafts turn on the left

0.261	0.258	0.249	0.251	0.247	0.256	0.250	0.247	0.255	0.243
0.252	0.250	0.253	0.247	0.251	0.243	0.258	0.251	0.245	0.250
0.248	0.252	0.254	0.250	0.247	0.253	0.251	0.246	0.249	0.252
0.247	0.250	0.253	0.247	0.249	0.253	0.246	0.251	0.249	0.253

Use the 0.01 Level of significance to the Null Hypothesis against the Alternative Hypothesis that there is a frequency Alternative pattern.

Soln: Null Hypothesis: Arrangement of sample value is Random

Alternate Hypothesis: There is a frequency alternating pattern

Level of significance: $0.01 = \alpha$

The Median of 40 measurements is 0.250 show that beign the following arrangement of value Above (or) Below 250.

a: Above Median = 0.250

b: Below Median = 0.250.

a a b a b a b a b a a b a b a a b b a a b a a
b b a b a a b b a b a b a.

Given:

$n_1 = 19$ (No of a's) ; $n_2 = 16$ (No of b's)

r = No of groups = 27 ; $r = 27$

formula:

$$Z = \frac{r - \mu_r}{\sigma_r} \quad ; Z \to \text{Calculated value.}$$

$$\mu_r \Rightarrow \frac{2n_1 n_2}{n_1 + n_2} + 1 \Rightarrow \frac{2(19)(16)}{19+16} + 1 \Rightarrow \mu_r = \frac{19(32)}{35} + 1$$

$$\boxed{\mu_r = 18.37}$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}} \Rightarrow \sigma_r = \sqrt{\frac{2(19)(16)\ (2(19)(16) - 19 - 16)}{(19+16)^2\ (19+16-1)}}$$

$$\sigma_r = \sqrt{\frac{608\ (573)}{35^2\ (34)}} \Rightarrow \boxed{\sigma_r = 2.89.}$$

$$Z = \frac{r - \mu_r}{\sigma_r} \Rightarrow Z = \frac{27 - 18.37}{2.89} \Rightarrow \boxed{Z = 2.99.}$$

↳ calculated value.

Table Value:

Level of significance : $\alpha = 0.01$ ⇒ $\boxed{\text{Table value} = 2.58}$

Conclusion:

Calculated value Z (2.99) > Table Value (2.58)

∴ The Hypothesis is Rejected.

Problem : 3

The following is the arrangement of defective, d, and non-defective, n, pieces produced in the given order by a certain machine:

nnnnn dddd nnnnnnnnnn dd nn dddd

Test for randomness at the 0.01 level of significance.

Soln:

1. Null Hypothesis : Arrangement is random

   Alternative Hypothesis: Arrangement is not random.

2. Level of significance: $\alpha = 0.01$

3. Criterion : Reject the null Hypothesis if $Z < -2.575$ (or) $Z > 2.575$, where Z is given by the above formula.

4. Calculations :

   Since $n_1 = 10$, $n_2 = 17$ and $r = 6$, we get

$$\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1 \Rightarrow \mu_r = \frac{2(10)(17)}{10+17} + 1 \Rightarrow \boxed{\mu_r = 13.59}$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} \; ; \; \sigma_r = \sqrt{\frac{2(10)(17)(2(10)(17) - 10 - 17)}{(10+17)^2(10+17-1)}}$$

$$\sigma_r = \sqrt{\frac{340(313)}{27^2(26)}} \; ; \; \boxed{\sigma_r = 2.37}$$

$$Z = \frac{r - \mu_r}{\sigma_r} = \frac{6 - 13.59}{2.37} \Rightarrow \boxed{Z = -3.20}$$

5. Decisions:

   Calculated value $Z = -3.20$ is less than -2.575, (table value at 1% level)

   $\therefore$ The Null Hypothesis must be Rejected.

We conclude that the arrangement is not random.

3. Mann - Whitney U-test (or) Wilcoxon Test :

→ Its a non-parametric test of null Hypothesis (2 possibilities are same):

formula:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where: n_1 - no of attributes of Group A

n_2 - no of attributes of Group B

R_1 - sum of all ranks of Group A

R_2 - sum of all ranks of Group B

Calculations:

$$Z = \frac{U - \left(\frac{n_1 n_2}{2}\right)}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

Where U can be any value from U_1 & U_2

Significance of table values:

1% = 2.58

5% = 1.96

If calculated value Z < table value ; Hypothesis is accepted.

If calculated value Z > table value ; Hypothesis is rejected.

Problem: 1

The method of instruction to apprentices is to be evaluated. A director assigns 15 randomly selected trainees to each of the two methods. Due to drop outs, 14 complete in batch 1 & 12 complete in batch 2. An achievement test was given to these successful candidates. Their scores as follows.

Method I: 70 90 82 64 86 77 84 79 82 89 73 81 83 66

Method II: 86 78 90 82 65 87 80 88 95 85 76 94

Test whether the two methods have significant difference in effectiveness. Use Mann-Whitney test for 5% significance.

Soln:

Method		Rank (overall)	Common Rank (R)	Rank.	
I	II			Rank × occurrance (R_1)	Rank × occurrance (R_2)
64	–	1	1	1	
–	65	2	2		2
66	–	3	3	3	
70	–	4	4	4	
73	–	5	5	5	
–	76	6	6		6
77	–	7	7	7	
–	78	8	8		8
79	–	9	9	9	
–	80	10	10		10
81	–	11	11	11	
82, 82	82	12, 13, 14	13	26	13
83	–	15	15	15	
84	–	16	16	16	
–	85	17	17		17
86	86	18, 19	18.5	18.5	
–	87	20	20		20

Method		Rank (overall)	Common Rank (R)	Rank.	
I	II			Rank x occurance (R1)	Rank x occurance (R2)
—	88	21	21		21
89	—	22	22	22	
90	90	23,24	23.5	23.5	23.5
—	94	25	25		25
—	95	26	26.		26.
				$R_1 = 161$	$R_2 = 190$

Given:

$n_1 = 14$; $n_2 = 12$; $R_1 = 161$; $R_2 = 190$

calculations:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_1 = 14(12) + \frac{14(14+1)}{2} - 161 \Rightarrow 168 + \frac{\overset{7}{\cancel{14}}(15)}{\cancel{2}} - 161$$

$$U_1 = 168 + 105 - 161 \Rightarrow \boxed{U_1 = 112}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2.$$

$$U_2 = 14(12) + \frac{\overset{6}{\cancel{12}}(12+1)}{\cancel{2}} - 190 \Rightarrow 168 + 6(13) - 190.$$

$$U_2 = 168 + 78 - 190 \Rightarrow \boxed{U_2 = 56}$$

$$Z = \frac{U - \left(\frac{n_1 n_2}{2}\right)}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \Rightarrow Z = \frac{56 - \left(\frac{14(\overset{6}{\cancel{12}})}{\cancel{2}}\right)}{\sqrt{\frac{14(\cancel{12})(14+12+1)}{\cancel{12}}}}$$

$$Z = \frac{56 - 14(6)}{\sqrt{14(27)}} \Rightarrow \frac{56-84}{\sqrt{378}} \Rightarrow \frac{-28}{19.44} = -1.44$$

$\boxed{Z = -1.44}$ (Sign Neglected)

Calculated value $Z = 1.44$

Table value at 5% LOS = 1.96.

Conclusion:

Calculated value "Z" (1.44) < Table Value (1.96)

∴ Hence Hypothesis is accepted.

∴ There is no significant difference b/w the 2 methods.

Problem: 2

The following are the two types of emergency flares on the basis of burning time (rounded to the nearest 10th of the minutes)

Brand A	14.9	11.3	13.2	16.6	17.0	14.1	15.4	13.0	16.9	
Brand B	15.2	19.8	14.7	18.3	16.2	21.2	18.9	12.2	15.3	19.4

Use the U-test at the 0.05 level of significance whether the 2 samples come from identical continuous populations are wether the average burning time of Brand A is less than Brand B flares.

Soln: Given:

Level of significance : $\alpha = 0.05$

n_1 = No of attributes in Brand "A" ; $n_1 = 9$

n_2 = No of attributes in Brand "B" ; $n_2 = 10$

BRAND		Common Rank	RANK	
A	B		Rank x occurance of A (R_1)	Rank x occurance of B (R_2)
11.3	—	1	1	
—	12.2	2		2.
13.0	—	3	3	
13.2	—	4	4	
14.1	—	5	5	
—	14.7	6		6
14.9	—	7	7	
—	15.2	8		8
—	15.3	9		9
15.4	—	10	10	
—	16.2.	11		11
16.6	—	12	12	
16.9	—	13	13	
17.0	—	14	14	
—	18.3	15		15
—	18.9	16		16
—	19.4	17		17
—	19.8	18		18
—	21.2.	19		19.
			$R_1 = 69$	$R_2 = 121$

Here

$n_1 = 9$; $n_2 = 10$; $R_1 = 69$; $R_2 = 121$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$U_1 = 9(10) + \frac{9(\cancel{10}^{5})}{\cancel{2}} - 69 \Rightarrow 90 + 45 - 69 \Rightarrow \boxed{U_1 = 66}$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2.$

$U_2 = 9(10) + \frac{\cancel{10}^{5}(11)}{\cancel{2}} - 121 \Rightarrow 90 + 55 - 121 \Rightarrow \boxed{U_2 = 24}$

$$Z = \frac{U - \left(\frac{n_1 n_2}{2}\right)}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} = \frac{24 - \left(\frac{9(10)}{2}\right)}{\sqrt{\frac{9(10)\ (9+10+1)}{12}}}$$

$$Z = \frac{24 - 45}{\sqrt{\frac{\cancel{90}^{30}\,(\cancel{20})^{5}}{\cancel{12}\ \cancel{3}}}} \Rightarrow Z = \frac{-21}{\sqrt{150}} \Rightarrow \boxed{Z = -1.71}$$

45
24
21

(sign Neglected)

∴ Calculated Value "Z" = 1.71

Table Value:

Level of significance : $\alpha = 0.05$ = 1.96.

Conclusion:

Table Value (1.96) > Calculated Value Z (1.71)

∴ Hypothesis is accepted.

∴ The Burning time of Brand A is less than Brand B.

4. Kruskal Wall's H-Test

→ It is a non-parametric test by ranks used to test wether sample originate from some distribution

→ Can have 2 (or) more samples.

formula:

$$H = \frac{12}{N(N+1)} \left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \cdots + \frac{R_n^2}{n_n} \right] - 3(N+1)$$

where

N = Total Count of sample value

R_1 = Rank sum of first sample.

R_2 = sum of Rank of second sample

n = no of Ranks in respective sample.

significance:-

$\vartheta = n-1$ = degree of freedom

where ϑ - significance.

n - no of samples provided for test

Table value to be calculated from chi-square Table.

If calculated value < table value.

∴ Hypothesis is accepted.

If calculated value > table value.

∴ Hypothesis is Rejected.

Problem 1:

Use the Kruskal-Wall's test to test for difference in mean among the three samples. If $\alpha = 0.01$. what is your conclusions?

Sample I :	95	97	99	98	99	99	99	94	95	98
Sample II :	104	102	102	105	99	102	111	103	100	103
Sample III :	119	130	132	136	141	172	145	150	144	135

Soln: Null Hypothesis : $H_0 = \mu_1 = \mu_2 = \mu_3$

Alternative Hypothesis : $H_1 = \mu_1, \mu_2, \mu_3$ are not equal

Level of significance $\alpha = 0.05$

Values	Common Ranks	Sample I Rank (R_1)	Sample II Rank (R_2)	Sample III Rank (R_3)
94	1	2.5	18	21
95, 95	2.5	4	14	22
97	4	9	14	23
98, 98	5.5	5.5	19	25
99, 99, 99, 99, 99	9	9	9	26
100	12	9	14	30
102, 102, 102	14	9	20	28
103, 103	16.5	1	16.5	29
104	18	2.5	12	27
105	19	5.5	16.5	24
111	20			
119	21	$R_1 = 57$	$R_2 = 153$	$R_3 = 255$
130	22			
132	23			

Values	Common Rank.
135	24
136	25
141	26
144	27
145	28
150	29
172	30

Given:

N = Total no of samples $\boxed{N = 30}$

$n_1 = 10$; $n_2 = 10$; $n_3 = 10$

$R_1 = 57$; $R_2 = 153$; $R_3 = 255$

formula:

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3}\right] - 3(N+1)$$

Calculation:

$$H = \frac{12}{30(31)}\left[\frac{57^2}{10} + \frac{153^2}{10} + \frac{255^2}{10}\right] - 3(30+1)$$

$$H = \frac{12}{930}\left[324.9 + 2340.9 + 6502.5\right] - 3(31)$$

$$H = \frac{12}{930}\left[9168.3\right] - 93 \Rightarrow \cancel{H = \frac{110019.6}{930} - 93}$$

$$H = 0.0129(9168.3) - 93 \Rightarrow 118.27 - 93$$

$\boxed{H = 25.27}$ → Calculated Value.

Table Value:

$\nu = n-1 \Rightarrow \nu = 3-1 \Rightarrow \nu = 2$. ; 1% ($\alpha = 0.01$)

$n=3$

Table value $= \psi^2_{0.01}(2) = 9.21$ (from Chi-Square Table)

$\boxed{\text{Table Value} = 9.21}$

Conclusion:

Calculated value = 25.27 ; Table value = 9.21

Calculated value > Table value

∴ Hypothesis is Rejected.

Problem: 2.

An experiment designed to compare three preventive methods against corrosion yielded the following maximum depths of pits (in thousands of an inch) in pieces of wire subject to the respective treatments:

Method A:	77	54	67	74	71	66	
Method B:	60	41	59	65	62	64	52
Method C:	49	52	69	47	56		

Use the 0.05 level of significance to test the null hypothesis that the three sample come from identical populations.

Soln:

Null Hypothesis: $H_0 = \mu_1 = \mu_2 = \mu_3$

Alternative Hypothesis: $H_1 = \mu_1, \mu_2, \mu_3$ are not equal

Level of significance $\alpha = 0.05$

formula:

$$H = \frac{12}{N(N+1)}\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3}\right] - 3(N+1)$$

Values	Common Rank.	Method A Rank (R_1)	Method B Rank (R_2)	Method C Rank (R_3)
41	1	18	9	3
47	2	6	1	4.5
49	3	14	8	15
52, 52	4.5	17	12	2
54	6	16	10	7
56	7	13	11	
59	8		4.5	
60	9			
62	10			
64	11	$R_1 = 84$	$R_2 = 55.5$	$R_3 = 31.5$
65	12			
66	13			
67	14			
69	15			
71	16			
74	17			
77	18			

Here:

$n_1 = 6$; $n_2 = 7$; $n_3 = 5$; $n = 3$

$R_1 = 84$; $R_2 = 55.5$; $R_3 = 31.5$

N = Total No. of Samples

$\boxed{N = 18}$

formula:

$$H = \frac{12}{N(N+1)} \left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3} \right] - 3(N+1)$$

$$H = \frac{12}{18(18+1)} \left[\frac{84^2}{6} + \frac{55.5^2}{7} + \frac{31.5^2}{5} \right] - 3(18+1)$$

$$H = \frac{12}{18(19)}\left[1176 + 440.036 + 198.45\right] - 3(19)$$

$$H = \frac{12}{342}\left[1814.486\right] - 57$$

$$H = 0.0351\left[1814.486\right] - 57$$

$$H = 63.688 - 57 \Rightarrow \boxed{H = 6.688}$$

Calculated Value "H" = 6.688.

Table Value:

Level of significance $\alpha = 0.05$; $\nu = n-1 \Rightarrow \nu = 3-1 = \boxed{\nu = 2}$

$\therefore$ Table Value = $\psi^2_{0.05}(2) = \boxed{5.991}$

Conclusion:

Calculated Value "H" (6.688) > Table Value (5.991)

$\therefore$ The Null Hypothesis must be Rejected.

We conclude that the preventive methods against corrosion are not equally effective.

Unit-5

Statistical Quality Control.

control charts for measurements- Control charts for attributes- Tolerance limits- Acceptance sampling.

control chart:-

A control chart provides a basis for deciding whether the variations in the output is due to random causes or due to assignable causes. It will assist us in making decisions whether to adjust the process or not.

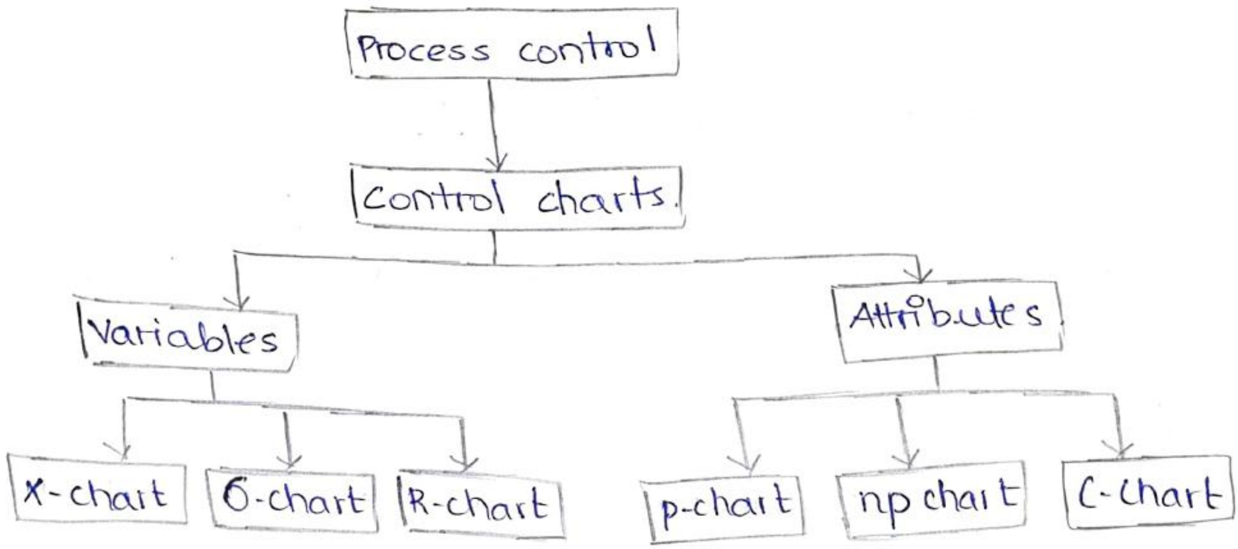
A control chart is designed to display successive measurements of a process with a centre line and control limits.

The control limits are above and below the center line and are equidistant from the centre line and are known as upper control limit (UCL) and lower control limit. (LCL)

The control charts helps us decide whether the process ~~prediction~~ of production is in control or not

Types of control charts:-

* Control charts for variables.
* Control charts for attributes.



Construction on $\bar{X}$-chart:-

~~Draw~~ Let $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$ be the means of these samples.

The mean of all these means is

$$X = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_n}{K}$$

The control limits are given by,

$$UCL = \bar{X} + 3.\, SE(\bar{X})$$

$$LCL = \bar{X} - 3.\, SE(\bar{X})$$

$$CL = \bar{X}$$

where, $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, σ being the SD of the production

If σ is not available the SD of the sampling Distribution of the mean can be taken as the best estimate of σ. In the case of small sample, the estimate of SE of $\bar{x}$ is $\frac{\sigma}{n-1}$. Alternatively in the case of small samples of size less than 20;

$$UCL = \bar{X} + A_2 \bar{R}$$

$$LCL = \bar{X} - A_2\bar{R}$$

$$CL = \bar{X}$$

Here $\bar{R}$ is the mean of the sample range $R_1, R_2, \ldots R_n$ obtained from k samples. The factor A_2 has to be determined from statistical tables when the sample size n is known.

Range Chart (R-chart)

For samples of size less than 20 the range provides a good estimate of σ. Hence to measure to measure the variance in the variable, range chart is used.

Construction of R-chart:-

Let $R_1, R_2, R_3, \ldots R_K$ be the values of the range in K samples. The mean of all these range is

$$\bar{R} = \frac{R_1 + R_2 + \ldots + R_K}{K}.$$

The control limits are given by,

$$LCL = D_3\bar{R}$$

$$UCL = D_4\bar{R}$$

The factors D_3 and D_4 are determined from statistical table for known sample size.

PROBLEMS BASED ON $\bar{X}$ AND $\bar{R}$ CHART :-

1. Given below are the values of sample mean $\bar{X}$ and sample range R for 10 samples, each of size 5. Draw the appropriate mean and range charts and comment on the state of control of the process.

Sample No :	1	2	3	4	5	6	7	8	9	10
Mean $\bar{X}_i$:	43	49	37	44	45	37	51	46	43	47
Range R_i :	5	6	5	7	7	4	8	6	4	6

Soln:-

$$\bar{X} = \frac{1}{N} \Sigma \bar{x}_i$$

$$= \frac{1}{10}(43+49+37+\dots+47) \qquad [\because N=10]$$

$$= 44.2$$

$$\bar{R} = \frac{1}{N}\Sigma R_i = \frac{1}{10}(5+6+5+\dots+6)$$

$$= 5.8$$

For sample size $n = 5$. (From the table of control chart)

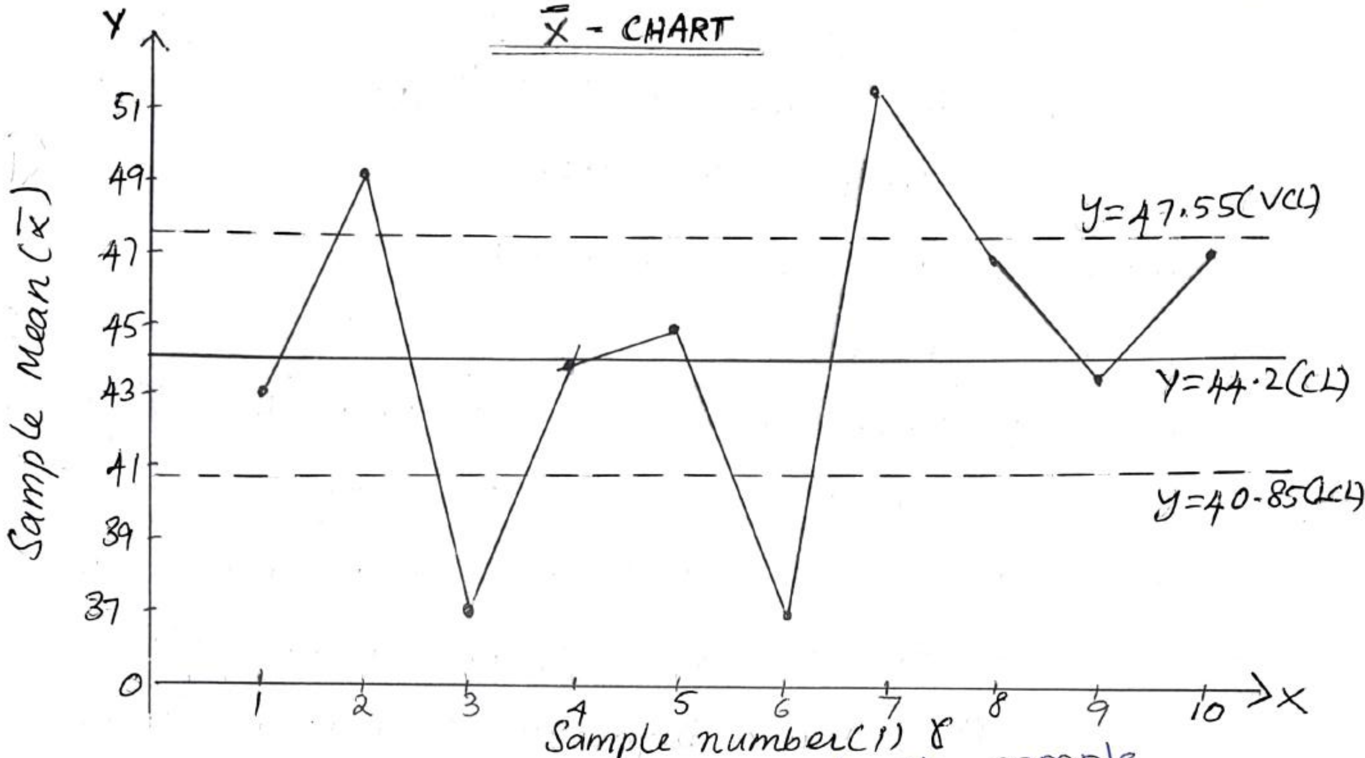
$A_2 = 0.577$, $D_3 = 0$ and $D_4 = 2.115$

i, Control limits for $\bar{X}$ chart:

CL (central line) = $\bar{X} = 44.2$

LCL = $\bar{X} - A_2\bar{R} = 44.2 - (0.577)(5.8) = 40.8534 \simeq 40.85$

UCL = $\bar{X} + A_2\bar{R} = 44.2 + (0.577)(5.8) = 47.5466 \simeq 47.55$

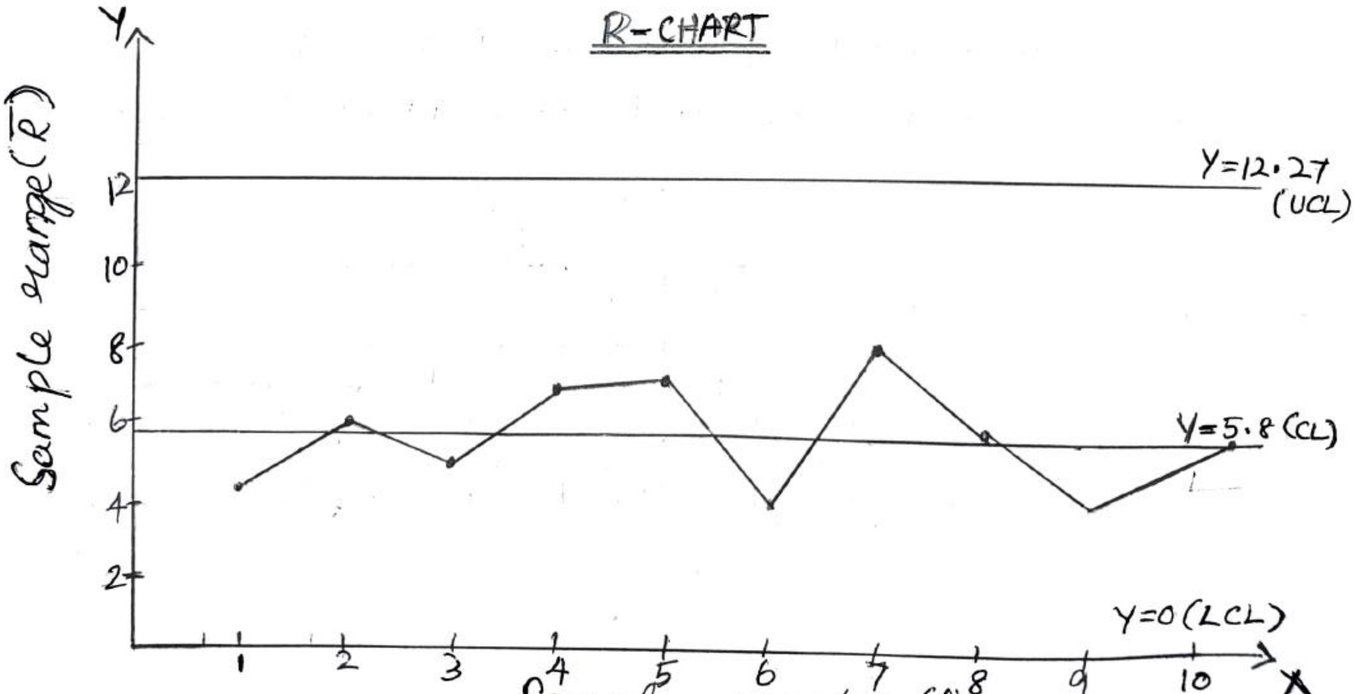


conclusion:- Since 2nd, 3rd, 6th and 7th sample means fall outside the control limits the statistical process is out of control according to $\bar{x}$-chart.

ii. Control limits for R-chart

$CL = \bar{R} = 5.8$; $LCL = D_3\bar{R} = 0$;

$UCL = D_4\bar{R} = (2.115)(5.8) = 12.267 \simeq 12.27$



Conclusion:- Since all the sample mean fall within the control limits the statistical process is under control according to R-chart.

Inference: From both $\bar{x}$ and R-chart, we see that a point in $\bar{x}$-chart lies outside control limits while all points in R-chart lies within control limits. Though the range variation is under control, we conclude that the process is out of statistical control.

Note :- i, If the process is to be under control, then all sample points in both $\bar{x}$ and R-chart must be within control limits.

ii, Eliminating the sample no. 8 which goes outside control limits, we can get new control limits to set up testing of quality

2. The following are the sample means and ranges for ten samples, each of size 5. Construct the control chart for mean and range and comment on the nature of control.

Sample No.	1	2	3	4	5	6	7	8	9	10
Mean:	12.8	13.1	13.5	12.9	13.2	14.1	12.1	15.5	13.9	14.2
Range:	2.1	3.1	3.9	2.1	1.9	3.0	2.5	2.8	2.5	2.0

Soln:-

$$\bar{X} = \frac{12.8 + 13.1 + 13.5 + \dots + 14.2}{10}$$

$$= \frac{135.3}{10} = 13.53$$

$$\bar{R} = \frac{\sum R_i}{N} = \frac{2.1 + 3.1 + \dots + 2.5 + 2.0}{10}$$

$$= \frac{25.9}{10} = 2.59$$

From the table of control charts constants, for sample size $n = 5$, $A_2 = 0.577$, $D_3 = 0$ and $D_4 = 2.115$

i, Control limits for $\bar{X}$-chart:

$$CL \rightarrow \bar{X} = 13.53$$

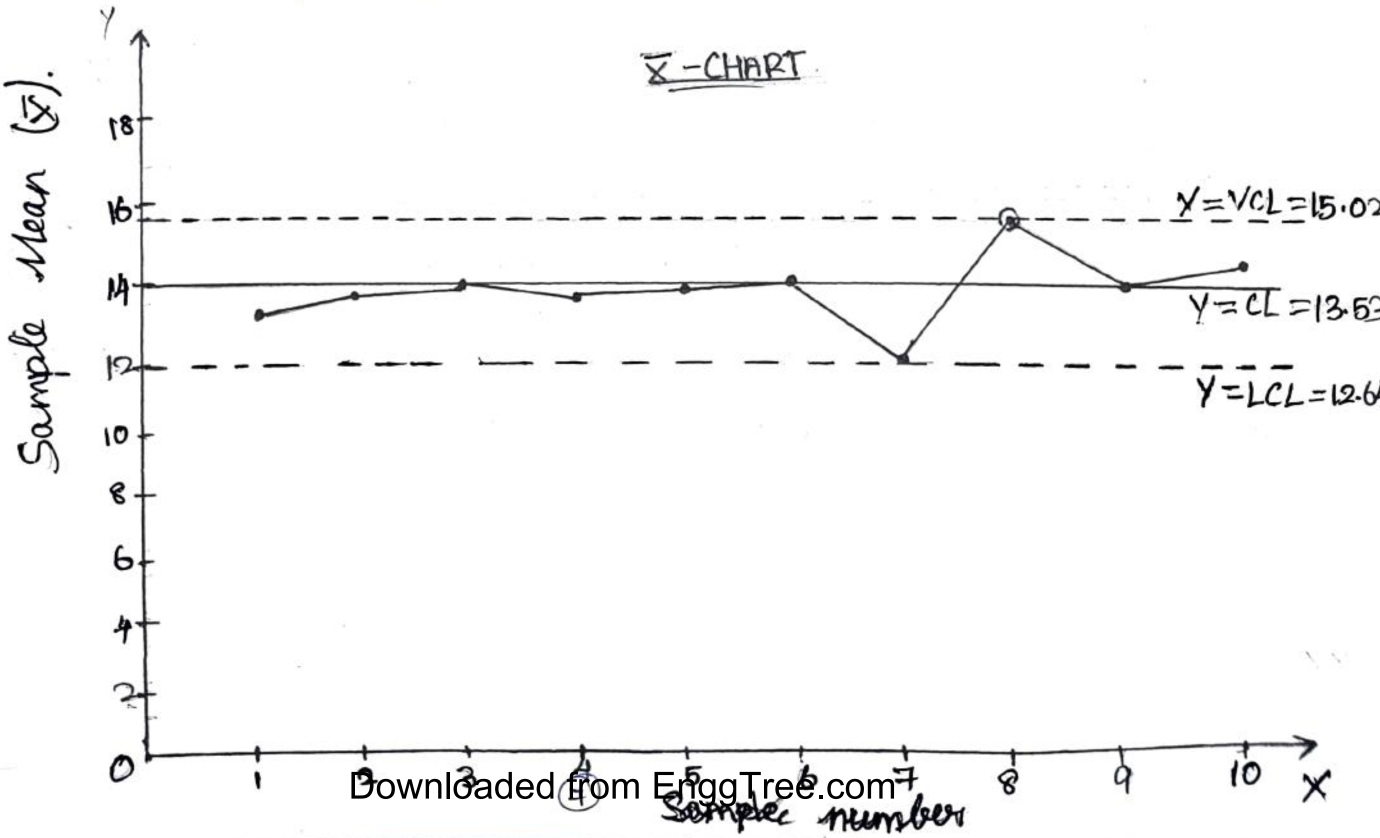
$$LCL = \bar{X} - A_2 \bar{R} = 13.53 - (0.577)(2.59)$$

$$= 12.03557 \simeq 12.04$$

$$UCL = \bar{X} + A_2 \bar{R}$$

$$= 13.53 + (0.577)(2.59) = 15.02443$$

$$\simeq 15.02$$



Conclusion:

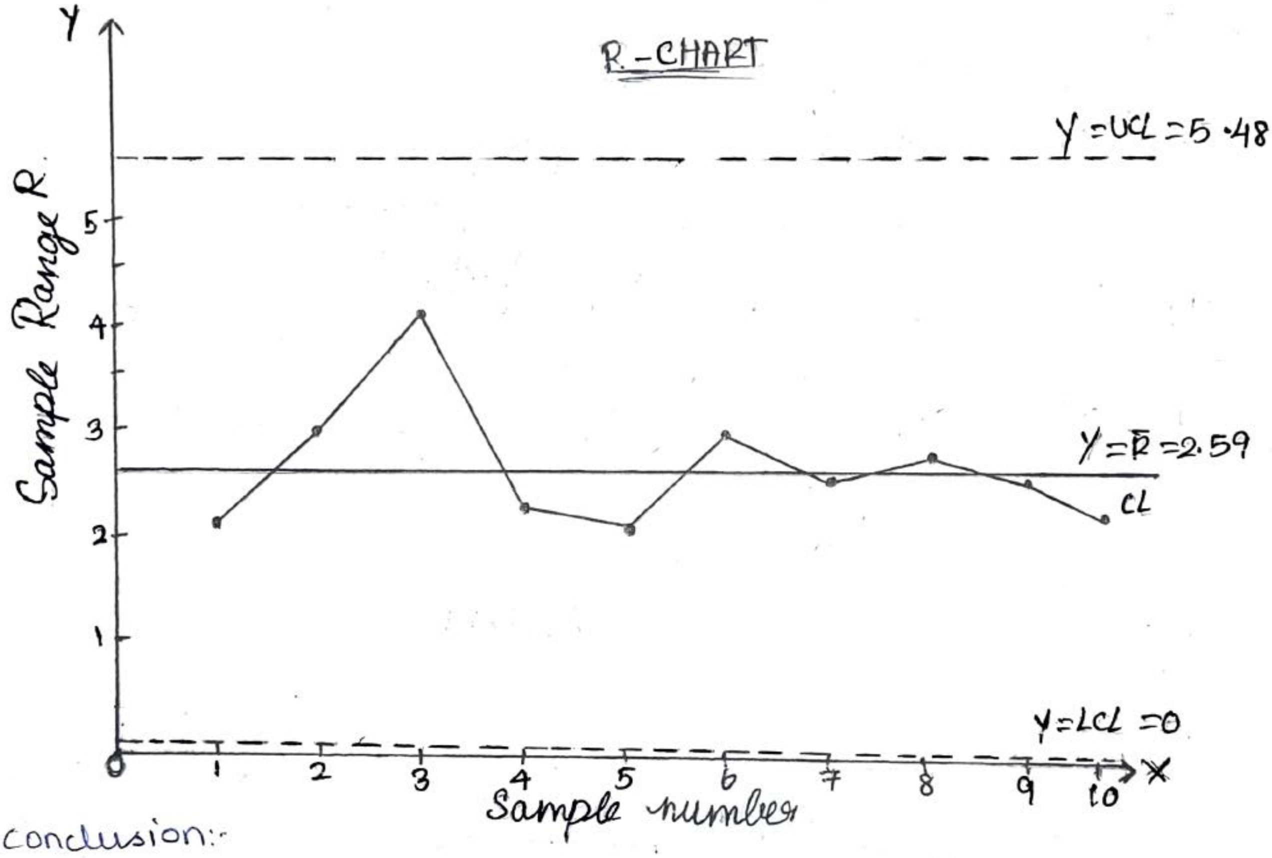
Since 8th sample mean fall outside the control limits the statistical process is out of control according to $\bar{X}$-chart.

ii, Control limits for R-chart:

$$UCL = D_4 \bar{R} = 2.115 \times 2.59 \simeq 5.48$$

$$LCL = D_3 \bar{R} = 0$$

$$CL \rightarrow \bar{R} = 2.59$$



Conclusion:-

Since all the sample mean fall within the control limits the statistical process is under control according to R-chart.

3. The following table gives the sample mean and range for 10 samples, each of size 6, in the production of certain component. Construct the control charts for mean and ~~average~~ range and comment on the nature of control.

Sample No:	1	2	3	4	5	6	7	8	9	10
Mean $\bar{x}$:	37.3	49.8	51.5	59.2	54.7	34.7	51.4	61.4	70.7	75.3
Range R:	9.5	12.8	10.0	9.1	7.8	5.8	14.5	2.8	3.7	8.0

Soln:-

$$\bar{\bar{X}} = \frac{\Sigma \bar{x}}{N} = \frac{37.3 + 49.8 + 51.5 + \cdots + 75.3}{10}$$

$$= \frac{546}{10} = 54.6$$

$$\bar{R} = \frac{\Sigma R}{N} = \frac{9.5 + 12.8 + \cdots + 8.0}{10}$$

$$= \frac{84.0}{10} = 8.4$$

From the table of control chart, for sample size of 6,

$A_2 = 0.483$, $D_3 = 0$, $D_4 = 2.004$

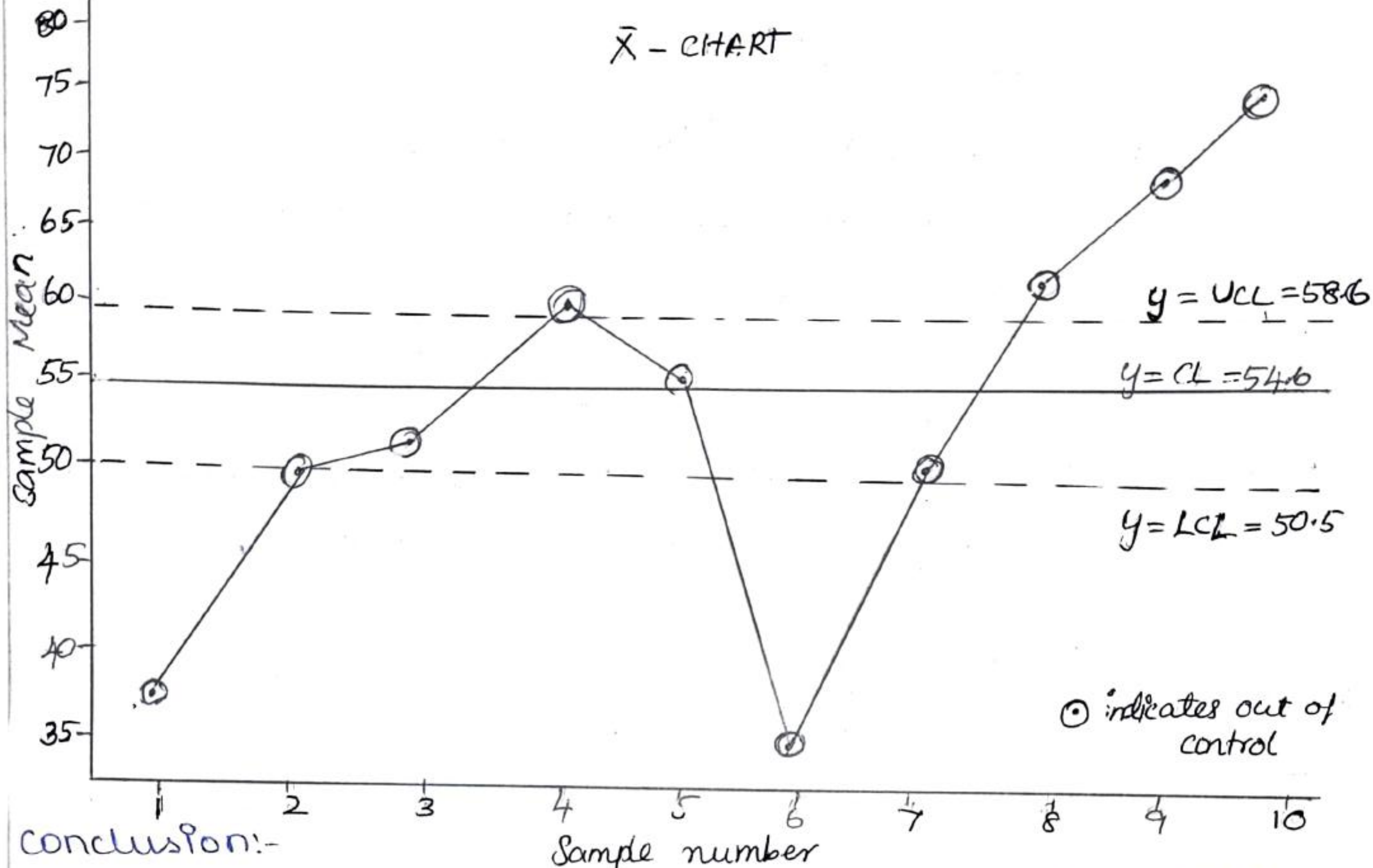
control limits of $\bar{x}$- chart

$$UCL = \bar{\bar{X}} + A_2\bar{R} = 54.6 + (0.483)(8.4)$$

$$= 58.657$$

$$LCL = \bar{\bar{X}} - A_2\bar{R} = 54.6 - (0.483)(8.4)$$

$$= 50.543$$

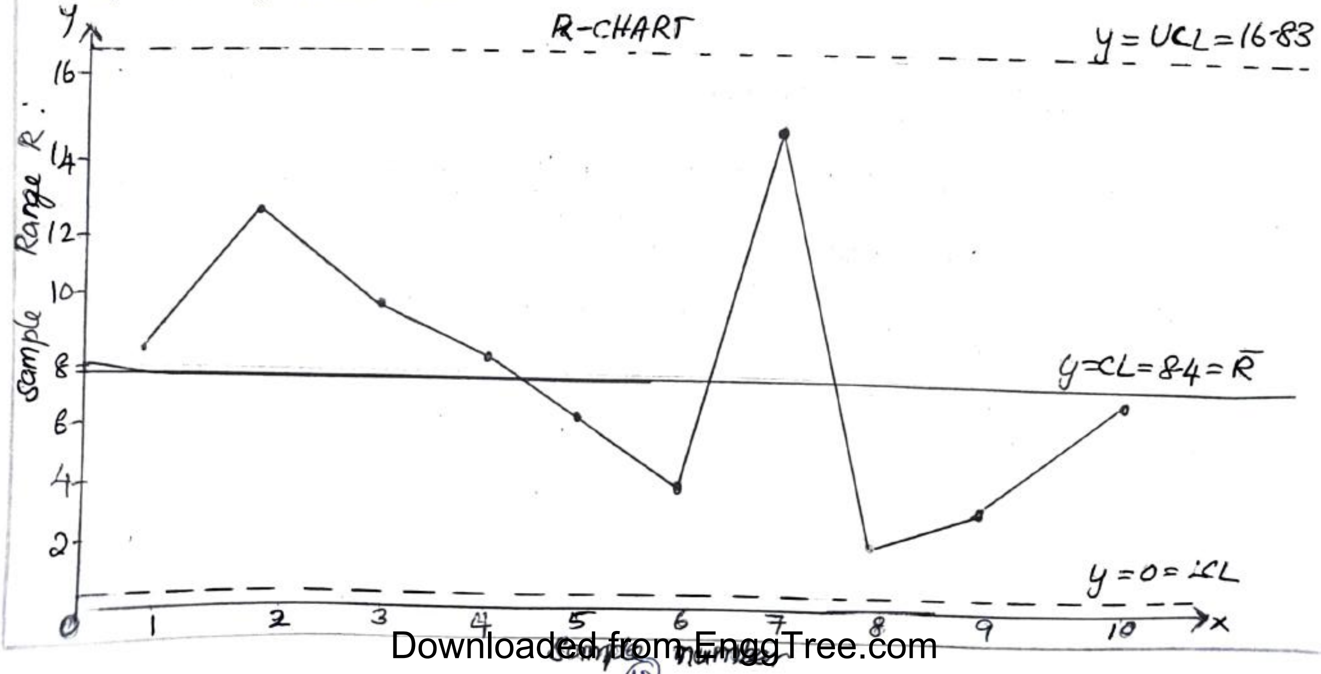


Conclusion:-

since 1st, 2nd, 4th, 6th, 8th, 9th, 10th sample means full outside the control limits the statistical process is out of control according to $\bar{x}$ chart.

control limits of R-chart:-

$\bar{R} = 8.4$; $UCL = D_4\bar{R} = 2.004 \times 8.4 = 16.834$; $LCL = D_3\bar{R} = 0$



Conclusion:-

since all the sample mean fall within the control lines the statistical process is under control according to R-chart.

Inference:- Though the sample points in R-chart lie within control limits, some of the sample points in $\bar{x}$-chart lie outside the control limits. Hence, we conclude that the process is out of control; corrective measures are necessary.

4. The following data give the measurements of 10 samples each of size 5 in the production process taken in an interval of 2 hours. calculate the sample means and ranges and draw the control charts for mean and range.

Sample Number	1	2	3	4	5	6	7	8	9	10
Observed Measurements X	49	50	50	48	47	52	49	55	53	54
	55	51	53	53	49	55	49	55	50	54
	54	53	48	51	50	47	49	50	54	52
	49	46	52	50	44	56	53	53	47	54
	53	50	47	53	45	50	45	57	51	56

Soln:-

We shall find $\bar{x}$ and R for each sample.

Sample Number	1	2	3	4	5	6	7	8	9	10
ΣX	260	250	250	255	235	260	245	270	255	270
$\bar{X}$	52	50	50	51	47	52	49	54	51	54
R	6	7	6	5	6	9	8	7	7	4

Soln:- $\bar{\bar{X}} = \frac{\Sigma \bar{X}}{N} = \frac{52+50+50+51+47+52+49+54+51+54}{10}$

$= \frac{510}{10} = 51$

$\bar{R} = \frac{\Sigma R}{N} = \frac{6+7+6+\ldots+7+7+4}{10} = 6.5$

From the table, for sample size of $n=5$,

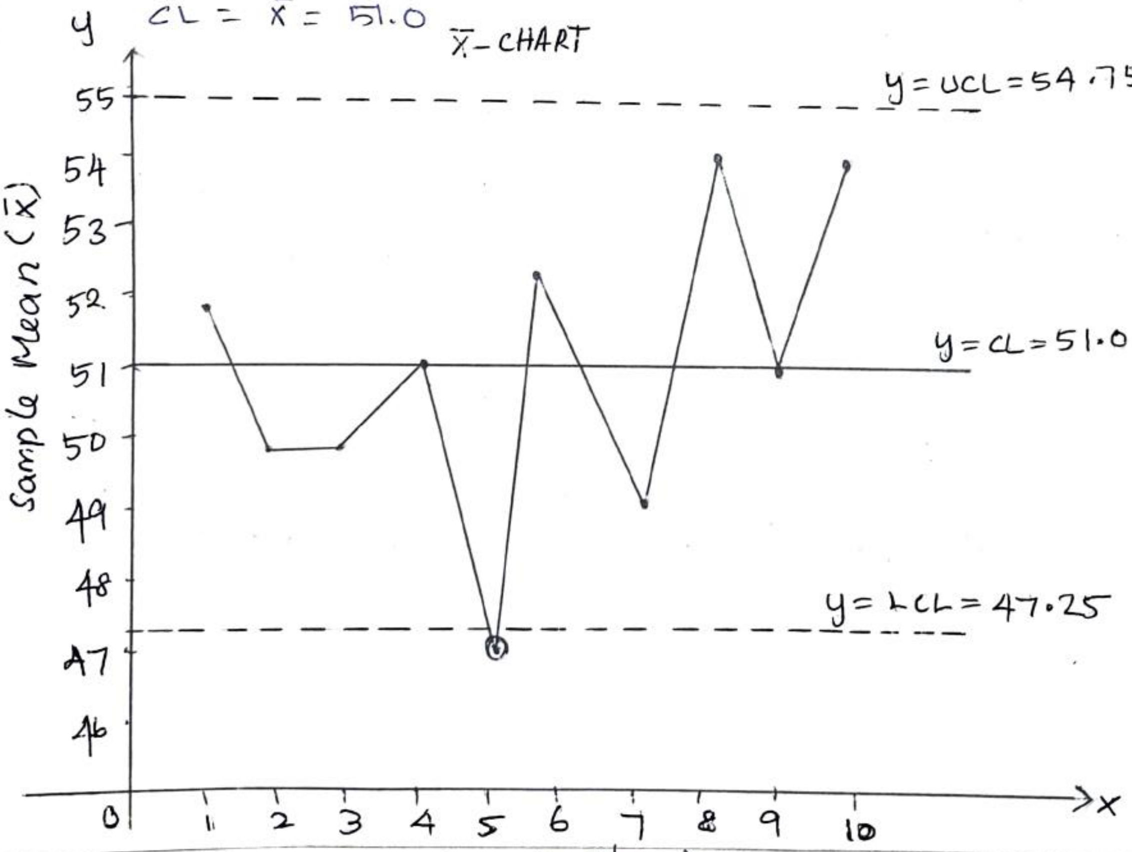
$A_2 = 0.577$, $D_3 = 0$, $D_4 = 2.115$

Control limits for X chart:

$UCL = \bar{\bar{X}} + A_2\bar{R} = 51.0 + (0.577)(6.5) = 54.7505$

$LCL = \bar{\bar{X}} - A_2\bar{R} = 51.0 - (0.577)(6.5) = 47.2495$

$CL = \bar{\bar{X}} = 51.0$



conclusion:

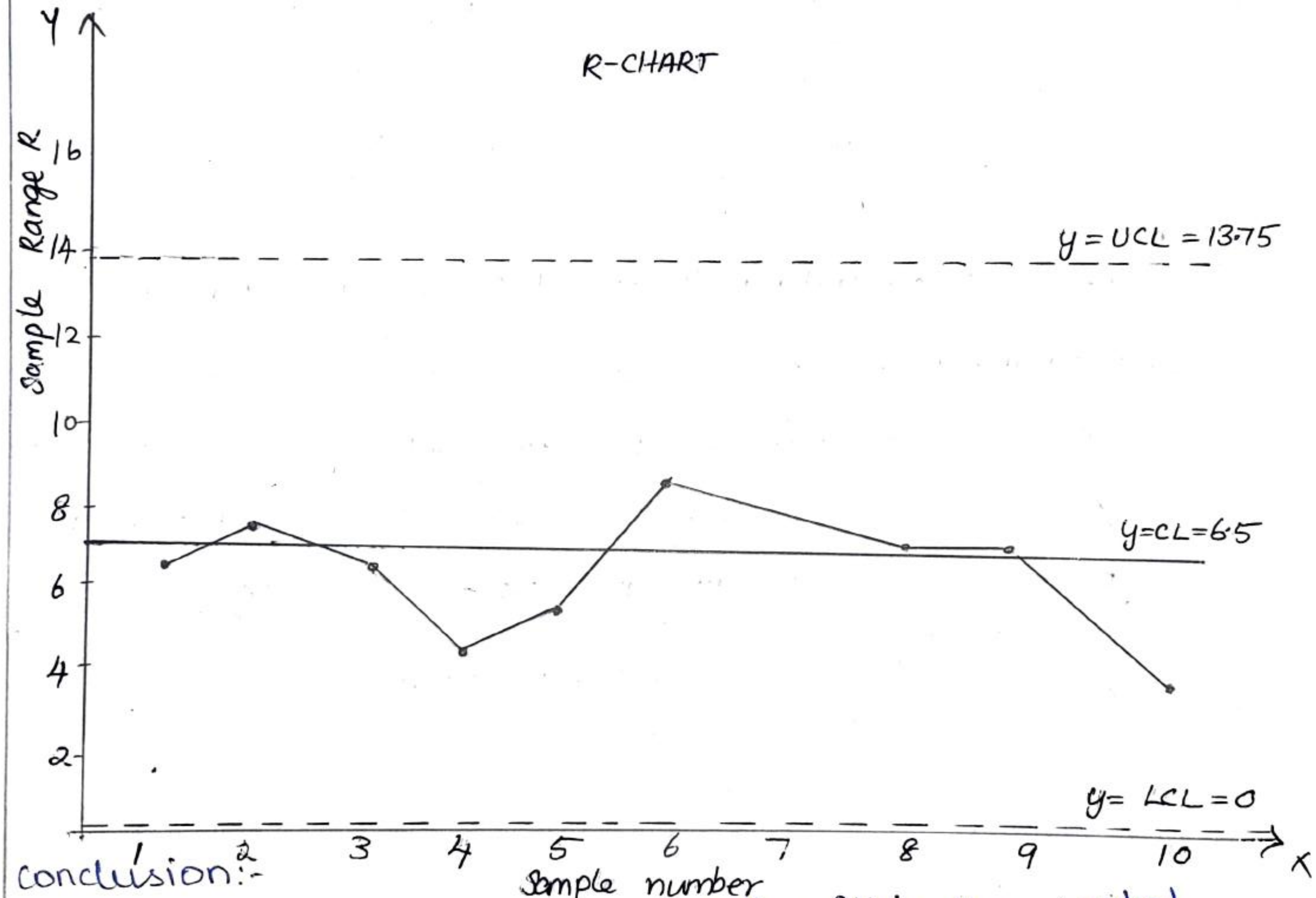
since 5th sample mean full outside the control limits the statistical process is out of control according to $\bar{x}$ chart.

Control limits for R-chart:

$UCL = D_4 \bar{R} = (2.115)(6.5) = 13.747$

$LCL = D_3 \bar{R} = 0$

$CL = \bar{R} = 6.5$



Conclusion:-

since all the sample mean fall within the control limits the statistical process is under control according to R-chart.

5. The table given below gives the measurements obtained in 10 samples. Construct control charts for mean and the range. Discuss the nature of control.

Sample Number	1	2	3	4	5	6	7	8	9	10
Measurements X	62	50	67	64	49	63	61	63	48	70
	68	58	70	62	98	75	71	72	79	52
	66	52	68	57	65	62	66	61	53	62
	68	58	56	62	64	58	69	53	61	50
	73	65	61	63	66	68	77	55	49	66
	68	66	66	74	64	55	53	57	56	75

Soln:-

We shall calculate $\bar{x}$ and R for each sample.

Sample Number	1	2	3	4	5	6	7	8	9	10
ΣX	405	349	388	382	406	381	397	361	346	395
X	67.5	58.2	64.7	63.7	67.7	63.5	66.2	60.1	57.7	65.8
R	11	16	14	17	49	20	24	19	31	25

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}}{n} = \frac{67.5 + 58.2 + 64.7 + \cdots + 65.8}{10}$$

$$= \frac{635.1}{10} = 63.51$$

$$\bar{R} = \frac{\Sigma R}{n} = \frac{226}{10} = 22.6$$

For sample size 6, $A_2 = 0.483$

$D_4 = 2.004$

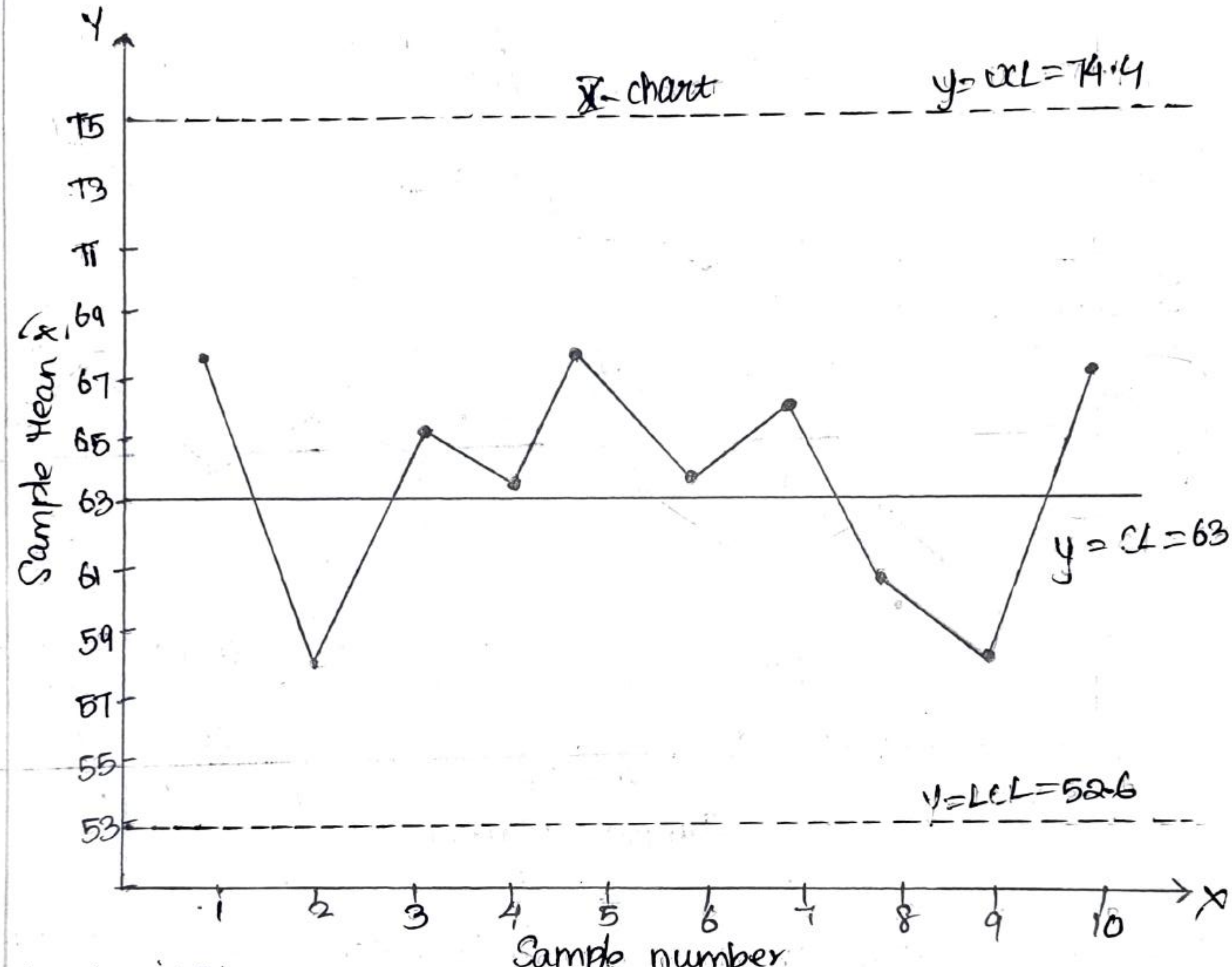
$D_3 = 0.$

Control limits for $\bar{x}$ - chart:

$$UCL = \bar{X} + A_2\bar{R} = 63.51 + (0.483)(22.6) = 74.43$$

$$LCL = \bar{X} - A_2\bar{R} = 63.61 - (0.483)(22.6) = 52.59$$

$$CL = \bar{X} = 63.51$$



Conclusion:

All the means of the sample lie between UCL and LCL

$52.59 < \text{all } \bar{X} < 74.43$

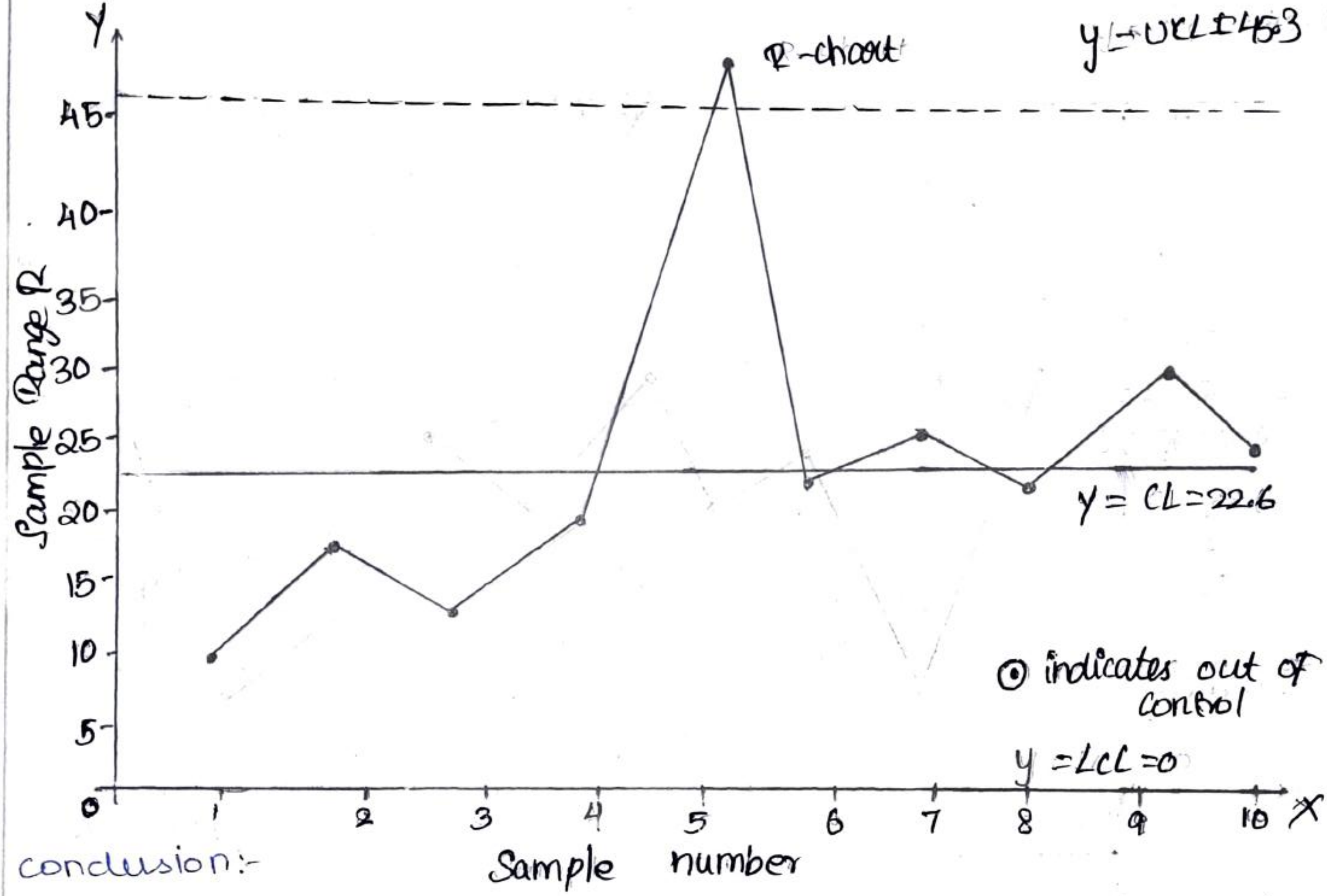
i.e. all sample points of means are falling within 3 sigma control limits. Hence, the process is in a state of statistical control as far as means are concerned.

Control limits for R-chart:

$UCL = D_4\bar{R} = 2.004 \times 22.6 = 45.29$

$LCL = D_3\bar{R} = 0$

$CL = \bar{R} = 22.6$



Conclusion:-

The value of R corresponding to sample no. 5, namely 49, lies outside the control limits. Hence the variability is out of control.

Inference:

The process is out of control due to R-chart

6. control on measurements of pitch diameter of thread in air-craft fitting is checked with 5 samples each containing 5 items at equal intervals of time. The measurements are given below. Construct $\bar{x}$ and R chart and state your inference from the charts.

Sample no.	Measurement				
1.	46	45	44	43	42
2.	41	41	44	42	40
3.	40	40	42	40	42
4.	42	43	43	42	45
5.	43	44	47	47	45

soln:-

For each sample, calculate $\bar{x}$ and R and tabulate:

sample no	Σx	$\bar{x}$	R
1	220	44.0	4
2	208	41.6	4
3	204	40.8	2
4	215	43.0	3
5	226	45.2	4

$$\bar{\bar{X}} = \frac{\Sigma \bar{x}}{5} = \frac{44+41.6+40.8+43.0+45.2}{5} = 42.92$$

$$\bar{R} = \frac{\Sigma R}{5} = \frac{17}{5} = 3.4$$

From table, for sample size 5 items,

$A_2 = 0.577$, $D_3 = 0$, $D_4 = 2.115$

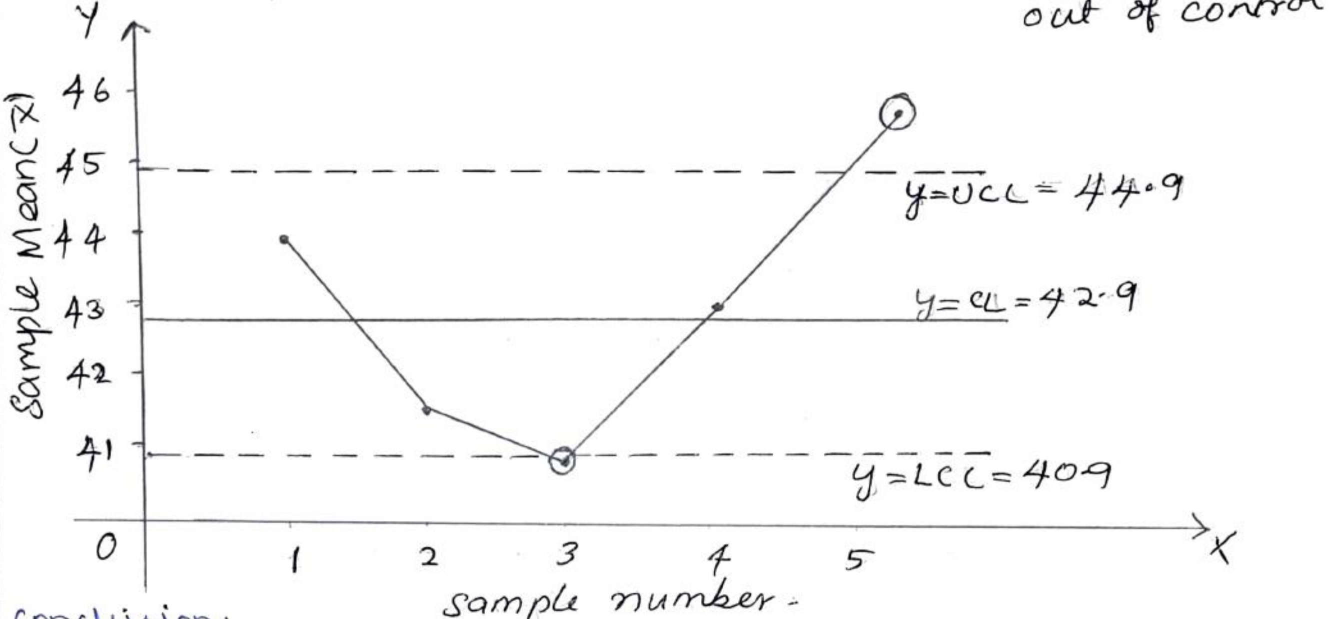
Limits for $\bar{x}$- chart:

$$UCL = \bar{\bar{X}} + A_2\bar{R} = 42.95 + (0.577)(3.4) = 44.88$$

$$LCL = \bar{\bar{X}} - A_2\bar{R} = 42.95 - (0.577)(3.4) = 40.96$$

$$CL = \bar{\bar{X}} = 42.92$$

⊙ indicates out of control



Conclusion:

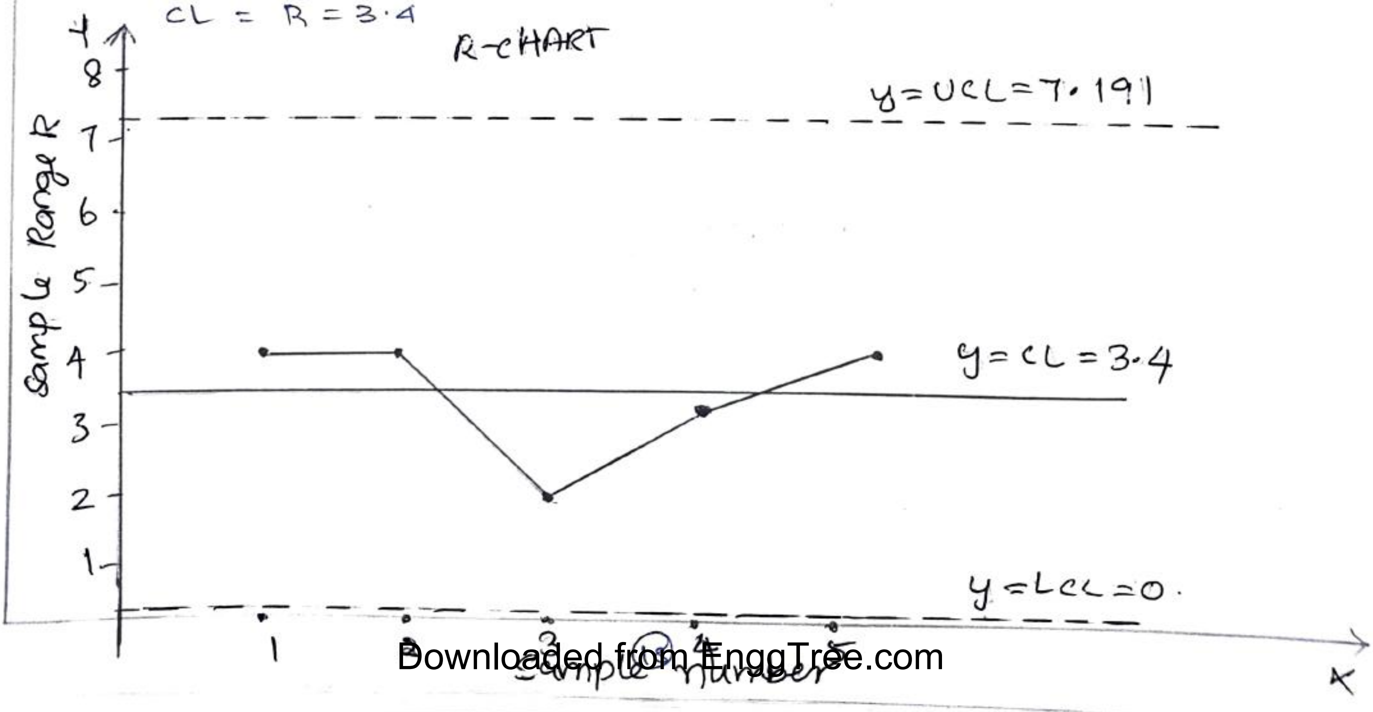
$\bar{x}_5 = 45.2 > UCL = 44.88$. All sample points do not lie between control limits. Hence, the process is out of control.

Limits for R-chart:-

$$UCL = D_4\bar{R} = (2.115)(3.4) = 7.191$$

$$LCL = D_3\bar{R} = 0$$

$$CL = \bar{R} = 3.4$$



Conclusion:

All sample points lie between control limits. Hence, the variability is under control. But process is out of control due to $\bar{x}$ - chart.

7. Construct an $\bar{x}$-R chart for the following data that give the heights of fragmentation bomb. Draw also the engineering specification tolerance limits of 0.830 ± 0.010 cm in the same graph. Infer your conclusion.

Group	Items.				
No.	1.	2.	3.	4.	5.
1.	0.831	0.829	0.836	0.840	0.826
2.	0.834	0.826	0.831	0.831	0.831
3.	0.836	0.826	0.831	0.822	0.816
4.	0.833	0.831	0.835	0.831	0.833
5.	0.830	0.831	0.831	0.833	0.820
6.	0.829	0.828	0.828	0.832	0.841
7.	0.835	0.833	0.829	0.830	0.841
8.	0.818	0.838	0.835	0.834	0.830
9.	0.841	0.831	0.831	0.833	0.832
10.	0.832	0.828	0.836	0.832	0.825

Soln:- In the given problem there are 10 sample groups of 5 each, that is N=10, n=5.

Group No.	1	2	3	4	5	6	7	8	9	10
sample Total	4.162	4.153	4.131	4.163	4.145	4.158	4.168	4.155	4.168	4.153
sample Range	0.014	0.008	0.020	0.004	0.013	0.013	0.012	0.020	0.010	0.011
sample Mean	.8324	.8306	.8262	.8326	.8290	.8316	.8336	.8310	.8336	.8306

From the statistical table, $A_2 = 0.577$

Grand Total = Total of sample totals = 41.556

$$\bar{X} = \frac{\text{Grand Total}}{(\text{Total no. of samples}) \,/\, (\text{sample size})}$$

$$= \frac{41.446}{(10)(5)} = \frac{41.446}{50}$$

$= 0.83112$

Total of sample ranges = 0.125

$$\bar{R} = \frac{\text{Total of sample ranges}}{\text{no. of samples}} = \frac{0.125}{10}$$

$= 0.0125$

Therefore, the control limits for $\bar{x}$-charts are:

$$\bar{X} \pm A_2 \bar{R}$$

i.e. $0.83112 \pm (0.577)(0.0125)$

i.e, 0.83112 ± 0.007212

UCL = 0.8383

LCL = 0.8239

The control limits for R-chart are given by

$$UCL = D_4\bar{R} \;;\; LCL = D_3\bar{R}$$

where D_3 and D_4 are constants taken from statistical table for n. Here $n = 5$, $\therefore$ $D_3 = 0$ and $D_4 = 2.115$

Hence UCL = $(2.115)(0.0125) = 0.0264$

LCL = 0

The process is under control.

Control chart for sample standard deviation or s-chart.

The standard deviation is an ideal measure of dispersion, a combination of control charts for the sample mean $\bar{x}$ and the sample s.D.

$\frac{s}{\sqrt{2n}}$, wher σ is the s.D of the population from which the sample is drawn.

$$\therefore P\left\{\sigma - \frac{3\sigma}{\sqrt{2n}} \le s \le \sigma + \frac{3\sigma}{\sqrt{2n}}\right\} = 0.9973$$

The lower and upper control limits are $\sigma - \frac{3\sigma}{\sqrt{2n}}$ and $\sigma + \frac{3\sigma}{\sqrt{2n}}$. Since σ is not known, it is estimated approximately by

$\bar{s} = \frac{1}{N}(s_1 + s_2 + \dots + s_N)$, where s_i the s.D of the i.th sample and N is the number of samples considered

$$\text{LCL for } s = \left(1 - \frac{3}{\sqrt{2n}}\right)\bar{s} \simeq B_3\bar{s} \text{ and}$$

$$\text{UCL for } s = \left(1 + \frac{3\sigma}{\sqrt{2n}}\right)\bar{s} \simeq B_4\bar{s}$$

The values of B_3 and B_4 can be read for various values of sample size n from the table of control chart constants.

If $\bar{x}$ values and s values only are given, then CL for $\bar{x} = \bar{\bar{x}}$, LCL for $\bar{x} = \bar{\bar{x}} - A_i\sqrt{\frac{n-1}{n}}\,\bar{s}$ and UCL for $\bar{x} = \bar{\bar{x}} + A_i\sqrt{\frac{n-1}{n}}\,\bar{s}$, when $n \le 25$.

8. The following data give the coded measurements of 10 samples each of size 5, drawn from a production process at intervals of 1 hour. Calculate the sample means and S.D's and draw the control charts for $\bar{x}$ and s.

Sample no.	1	2	3	4	5	6	7	8	9	10
Coded Measurements (x)	9	10	10	8	7	12	9	15	10	16
	15	11	13	13	9	15	9	15	13	14
	14	13	8	11	10	7	9	10	14	12
	9	6	12	10	4	16	13	13	7	14
	13	10	7	13	5	10	5	17	11	14

Soln:-

Sample no.	1	2	3	4	5	6	7	8	9	10
Σx	60	50	50	55	35	60	45	70	55	70
$\bar{x}$	12	10	10	11	7	12	9	14	11	14
$\Sigma(x-\bar{x})^2$	32	26	26	18	26	54	32	28	30	8
$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$	2.5	2.3	2.3	1.9	2.3	3.3	2.5	2.4	2.4	1.3

$$\bar{\bar{x}} = \frac{1}{N}\Sigma\bar{x}_i = \frac{1}{10} \times (12+10+10+\dots+14) = \frac{110}{10} = 11$$

$$\bar{s} = \frac{1}{N}\Sigma s_i = \frac{1}{10} \times (2.5+2.3+\dots+1.3) = \frac{23.2}{10} = 2.32$$

From the table, for sample size n=5,

$A_1 = 1.596$, $B_3 = 0$; $B_4 = 2.089$

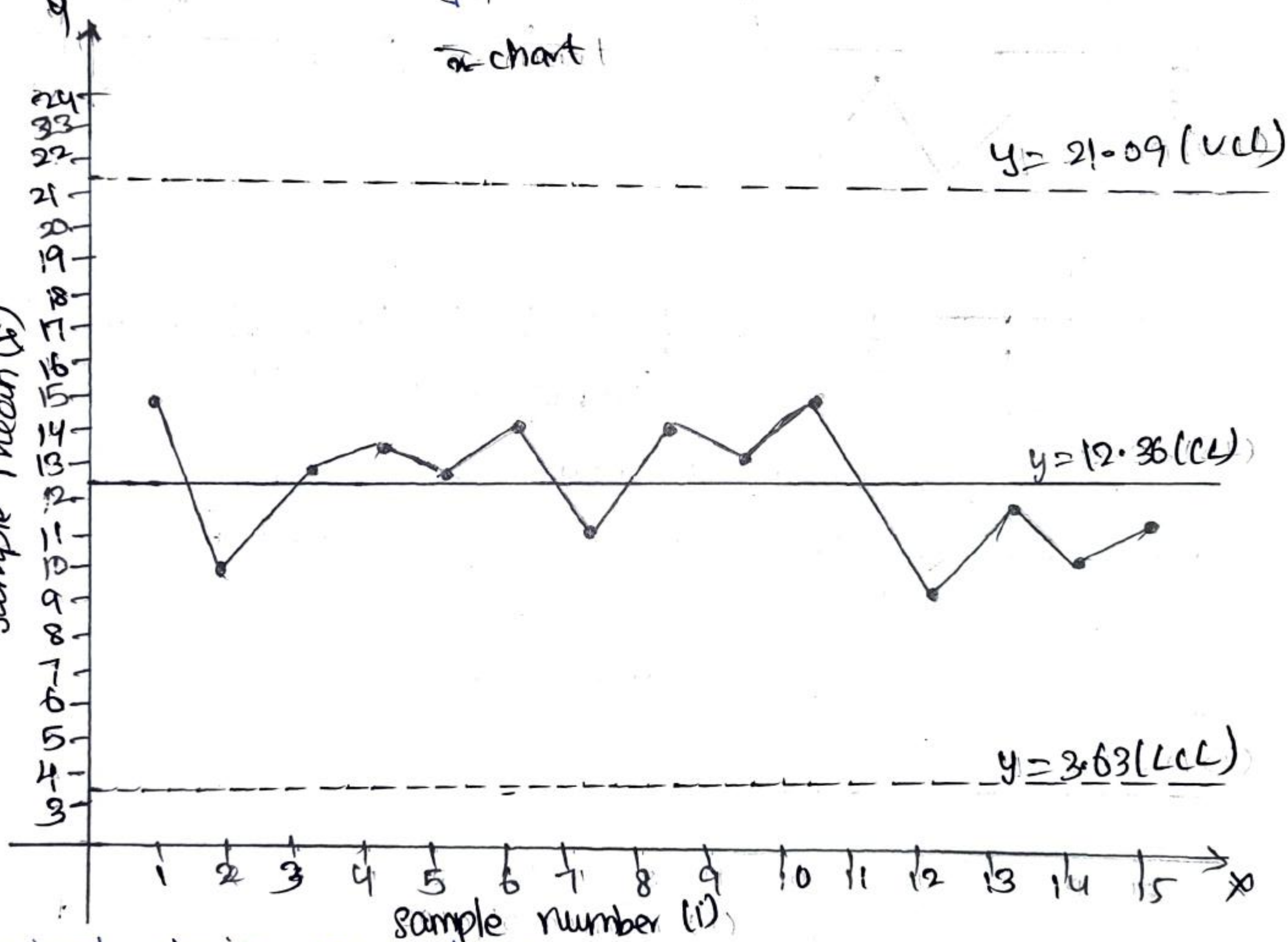
control limits for $\bar{x}$-chart:

$CL = \bar{\bar{x}} = 11$; $LCL = \bar{\bar{x}} - A_1 \sqrt{\frac{n}{n-1}}\ \bar{s}$

$= 11 - 1.596\sqrt{\frac{5}{4}} \times 2.32 = 6.86$

$UCL = \bar{\bar{x}} + A_1 \sqrt{\frac{n}{n-1}}\ \bar{s}$

$= 11 + 1.596\sqrt{\frac{5}{4}} \times 2.32 = 15.14$

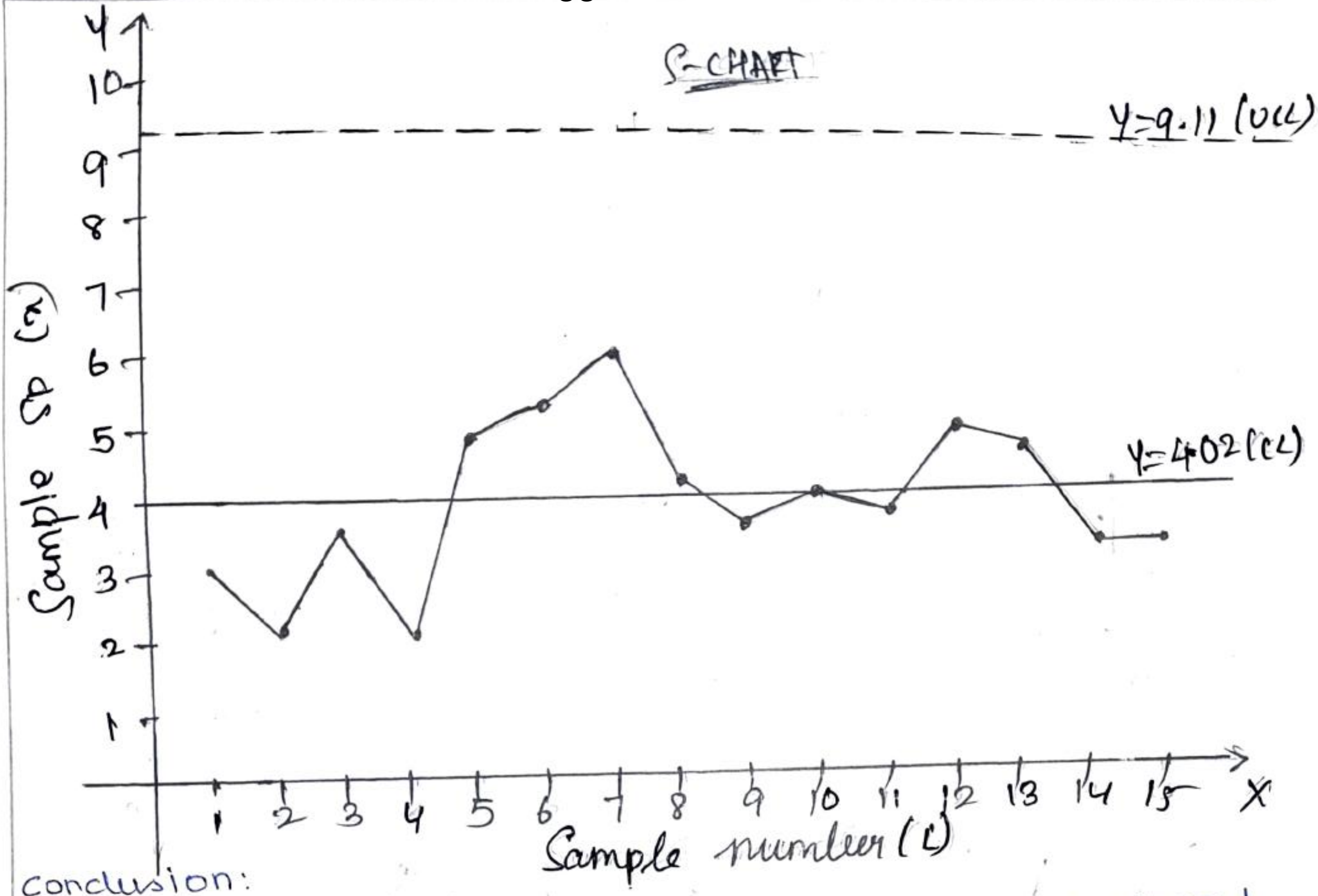


control limits for s-chart

$CL = \bar{s} = 2.32$

$LCL = B_2 \bar{s} = 0$; $UCL = B_4 \bar{s} = (2.089)(2.32)$

$= 4.85$



Conclusion:

The sample mean ($\bar{x}$) values lie between 6.86 and 15.14 and the given S.D (s) value between 0 and 4.85

Hence the process is under control with respect to average and variability.

9. The values of sample mean $\bar{x}$ and sample s.d s for 15 samples, each of size 4, drawn from a production process are given below Draw the appropriate control charts for the process average and process variability comment on the state of control.

Sample No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Mean	15.0	10.0	12.5	13.0	12.5	13.0	13.5	11.5	13.5	13.0	14.5	9.5	12.0	10.5	11.5
S.D	3.1	2.4	3.6	2.3	5.2	5.4	6.2	4.3	3.4	4.1	3.9	5.1	4.7	3.3	3.3

soln:-

$$\bar{\bar{x}} = \frac{1}{N} \sum \bar{x}_i = \frac{1}{15} \times 185.5 = 12.36$$

$$\bar{s} = \frac{1}{N} \sum s_i = \frac{1}{15} \times (60.3) = 4.02$$

From the table, for sample size $n = 4$,

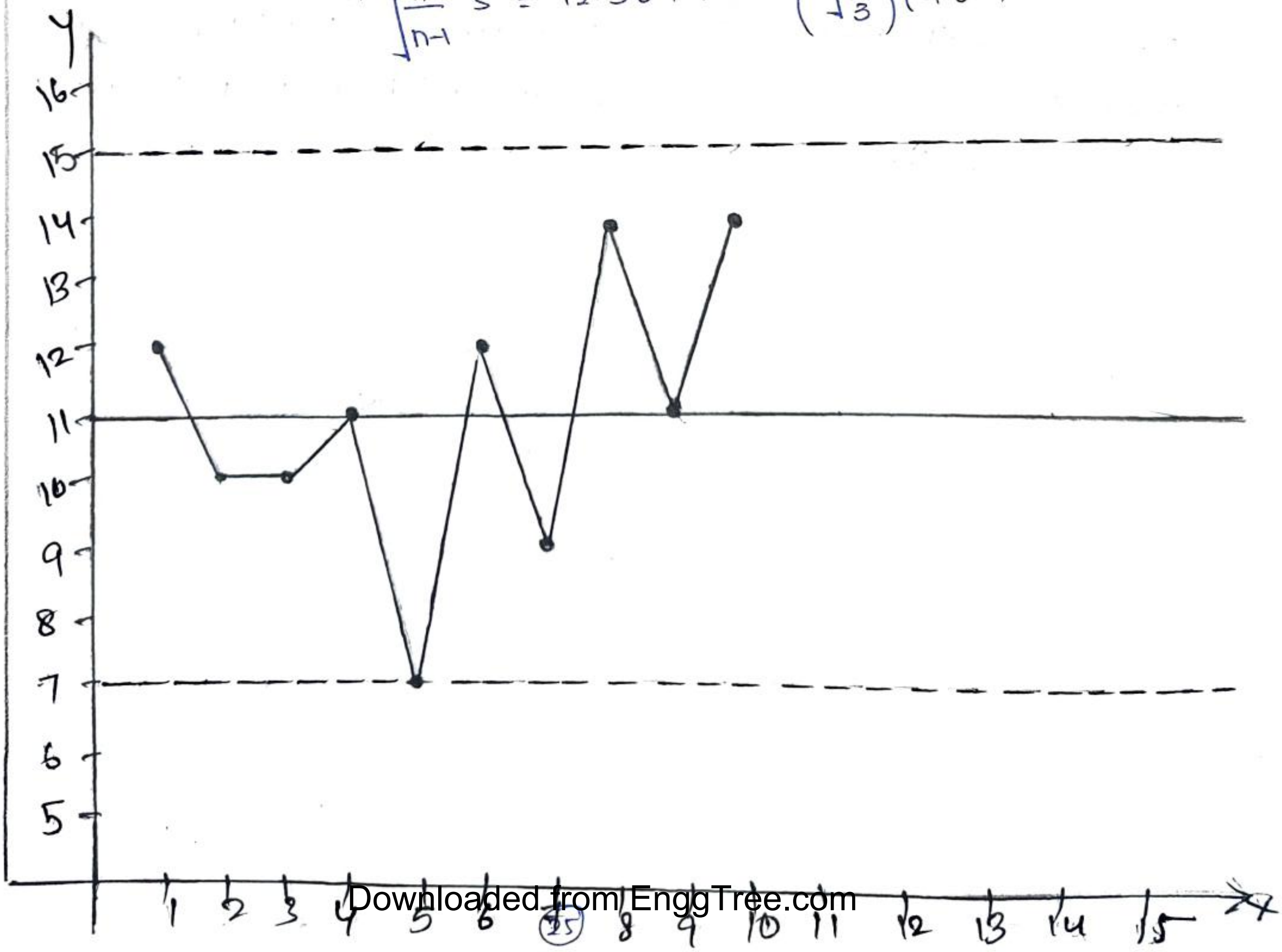
$A_1 = 1.880$, $B_3 = 0$, $B_4 = 2.266$

Control limits for $\bar{x}$-chart

$CL = \bar{\bar{x}} = 12.36$

$$LCL = \bar{\bar{x}} - A_1 \sqrt{\frac{n}{n-1}}\, \bar{s} = 12.36 - 1.880\left(\sqrt{\frac{4}{3}}\right)(4.02) = 3.63$$

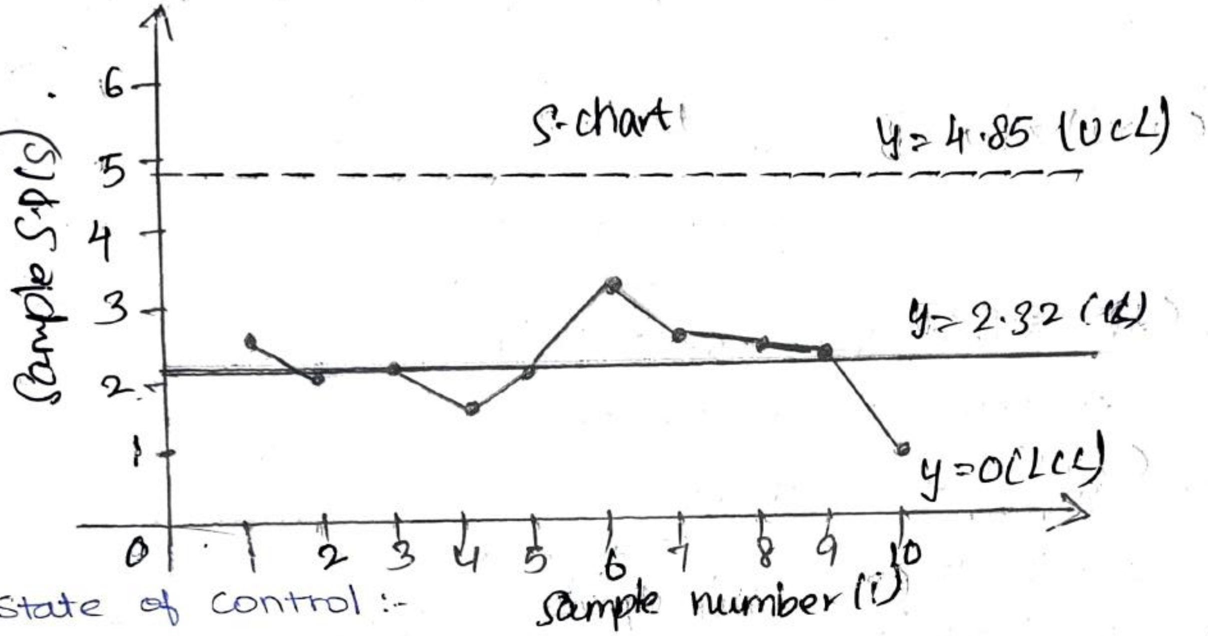
$$UCL = \bar{\bar{x}} + A_1 \sqrt{\frac{n}{n-1}}\, \bar{s} = 12.36 + 1.880\left(\sqrt{\frac{4}{3}}\right)(4.02) = 21.09$$



Control limits for s-chart.

$CL = \bar{s} = 4.02$; $LCL = B_3 \bar{s} = 0$;

$UCL = B_4 \bar{S} = (2.266)(4.02) = 9.11$

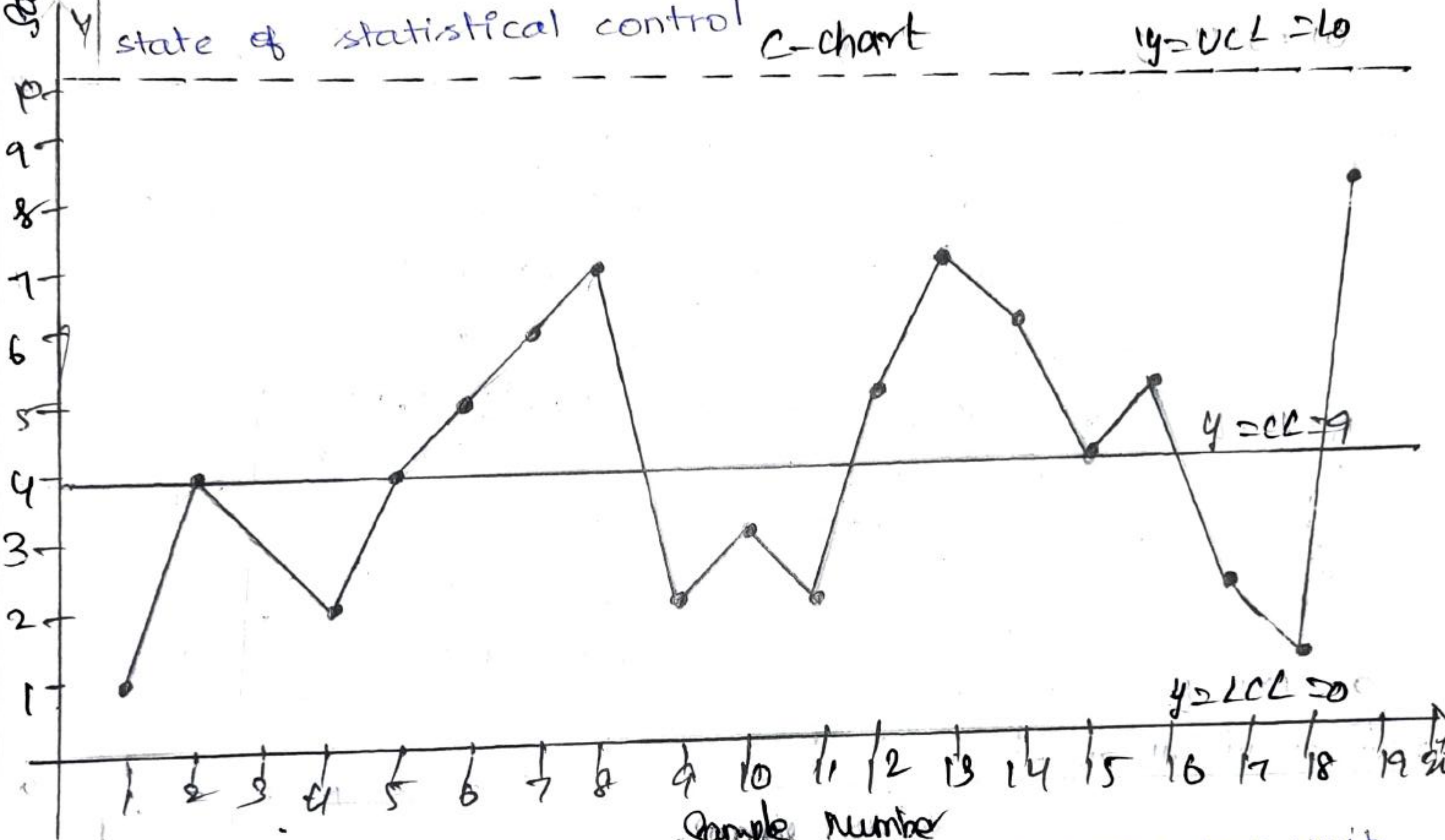


State of control :-

Even before drawing the control charts, we observe that the given sample mean value lies between 3.63 and 21.09 and that the given S.D values fall within 0 and 9.11. Hence the process is under control with respect to average and variability

Control charts for attributes:-

10. 20 pieces of cloth out of different rolls contained respectively 1, 4, 3, 2, 4, 5, 6, 7, 2, 3, 2, 5, 7, 6, 4, 5, 2, 1, 3 and 8 imperfections. Ascertain whether the process in a state of statistical control



soln:-

Let c denote the number of imperfections per unit

$$\bar{c} = \frac{\text{Total no. of defects}}{\text{Total sample inspected}} = \frac{\Sigma c}{n}$$

$$= \frac{1+4+3+2+\cdots+1+3+8}{20} = \frac{80}{20}$$

$\bar{c} = 4$

$CL(c) = 4$

$\sigma = \sqrt{\bar{c}} = \sqrt{4} = 2$

$UCL = \bar{c} + 3\sqrt{\bar{c}} = 4 + 3 \times 2 = 10$

$LCL = \bar{c} - 3\sqrt{\bar{c}} = 4 - 6 = -2$

Since LCL is negative take LCL = 0.

∵ all the values of c in the problem lie between LCL = 0 and UCL = 10 the process is under control

11. A textile unit produces special cloths and packs them in rolls. The number of defects found in 20 rolls are given below. Find whether the process is under control.

Defects in 20 rolls: 12, 14, 7, 6, 10, 10, 10, 11, 12, 5, 18, 12, 4, 4, 9, 21, 14, 8, 9, 13, 21

Soln:-

Let c denote the number of defects:

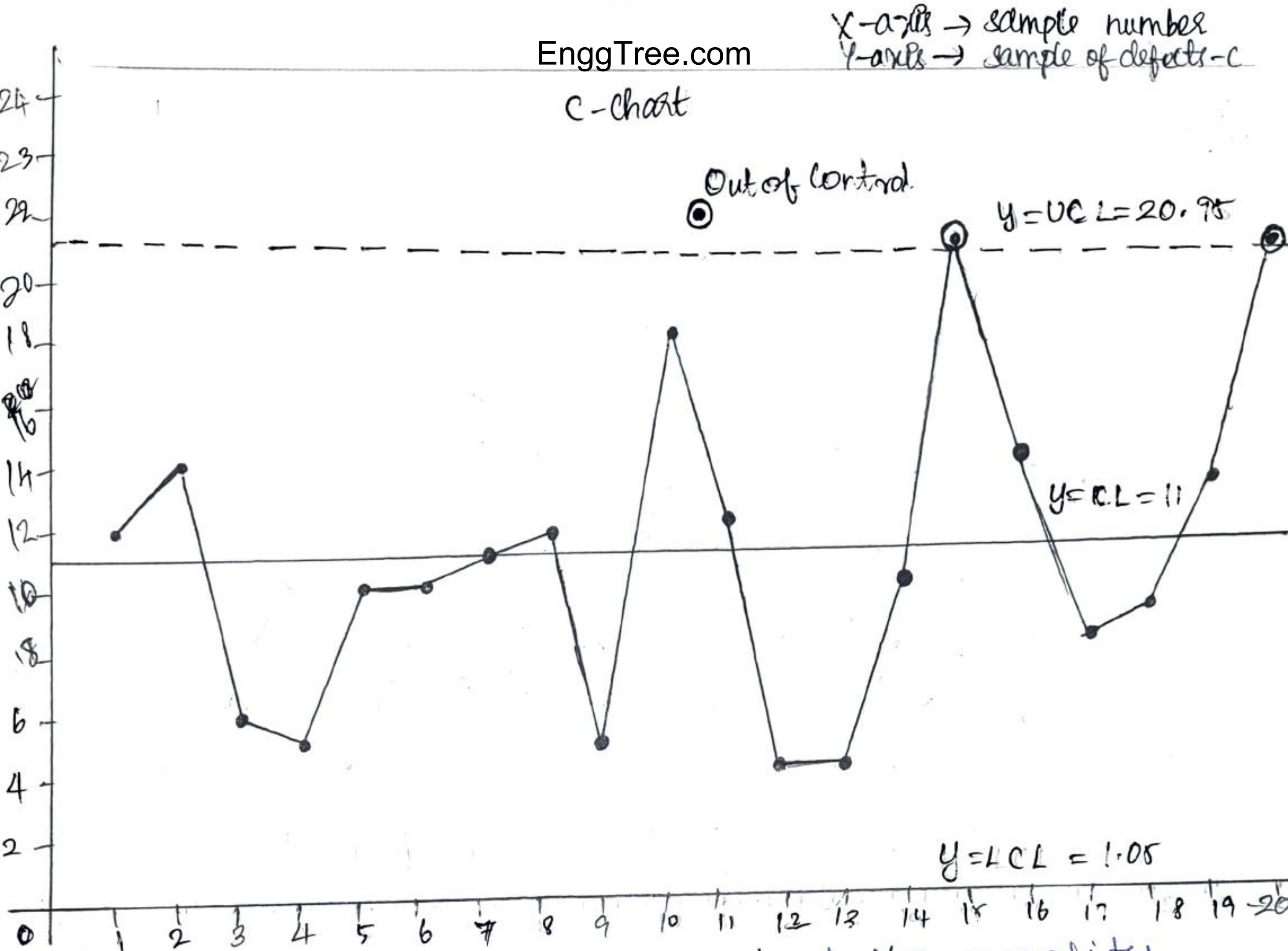
$$\bar{c} = \frac{\Sigma c}{n} = \frac{12+14+7+6+\dots+9+13+21}{20} = \frac{220}{20} = 11$$

$UCL = \bar{c} + 3\sqrt{\bar{c}} = 11 + 3\sqrt{11} = 20.95$

$LCL = \bar{c} - 3\sqrt{\bar{c}} = 11 - 3\sqrt{11} = 1.05$

$CL = c = 11.$

The inspection of values of c, we find two values of c, namely 21, 21 are greater than UCL = 20.95. These two values of c lie outside the control limits. Hence the process is out of control.



12. 15 tape recorders were examined for quality control test. The number of defects in each tape recorder is recorded below. Draw the appropriate control chart and comment on the state of control.

Unit no (i):	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of defects (c)	2	4	3	1	1	2	5	3	6	7	3	1	4	2	1

Soln:- The number of defects per sample containing only one item is given,

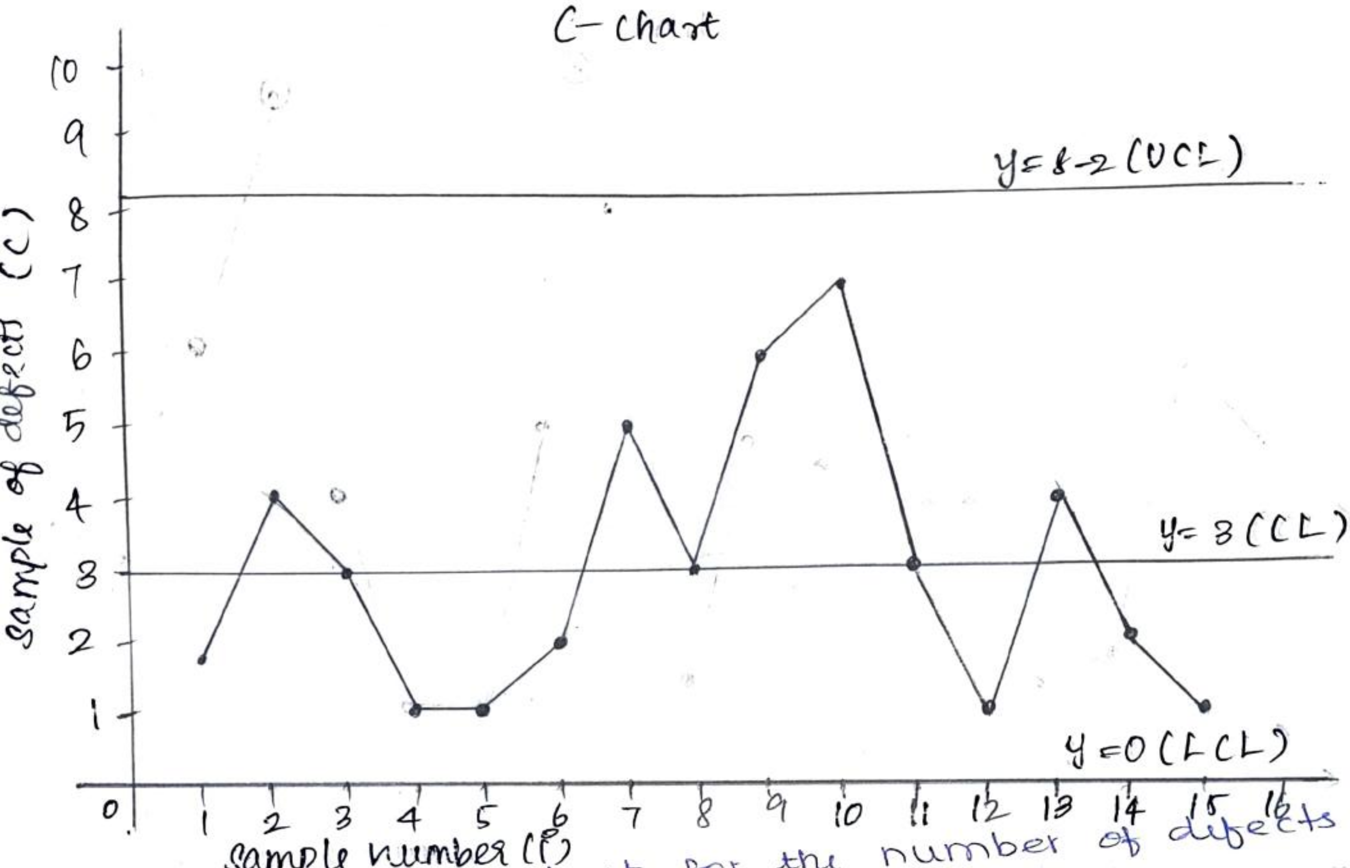
$$\therefore \bar{c} = \frac{1}{N} \Sigma c_i = \frac{1}{15}(2+4+\cdots+2+1) = \frac{45}{15} = 3$$

$CL = \bar{c} = 3$; $LCL = \bar{c} - 3\sqrt{c} = 3 - 3\sqrt{3} = -2.20$.

Since LCL cannot be -ve

$LCL = 0$.

$UCL = \bar{c} + 3\sqrt{c} = 3 + 3\sqrt{c} = 8.20$



13. Construct a control chart for the number of defects from the following data which give the number of defects in 15 pieces of cloth of equal length when inspected in a textile mill and find the nature of the process. Number of defects:

3, 4, 2, 7, 9, 6, 5, 4, 8, 10, 5, 8, 7, 7, 5.

soln:- Let c denote the number of defects in each piece

$$\bar{c} = \frac{\Sigma c}{n} = \frac{90}{15} = 6$$

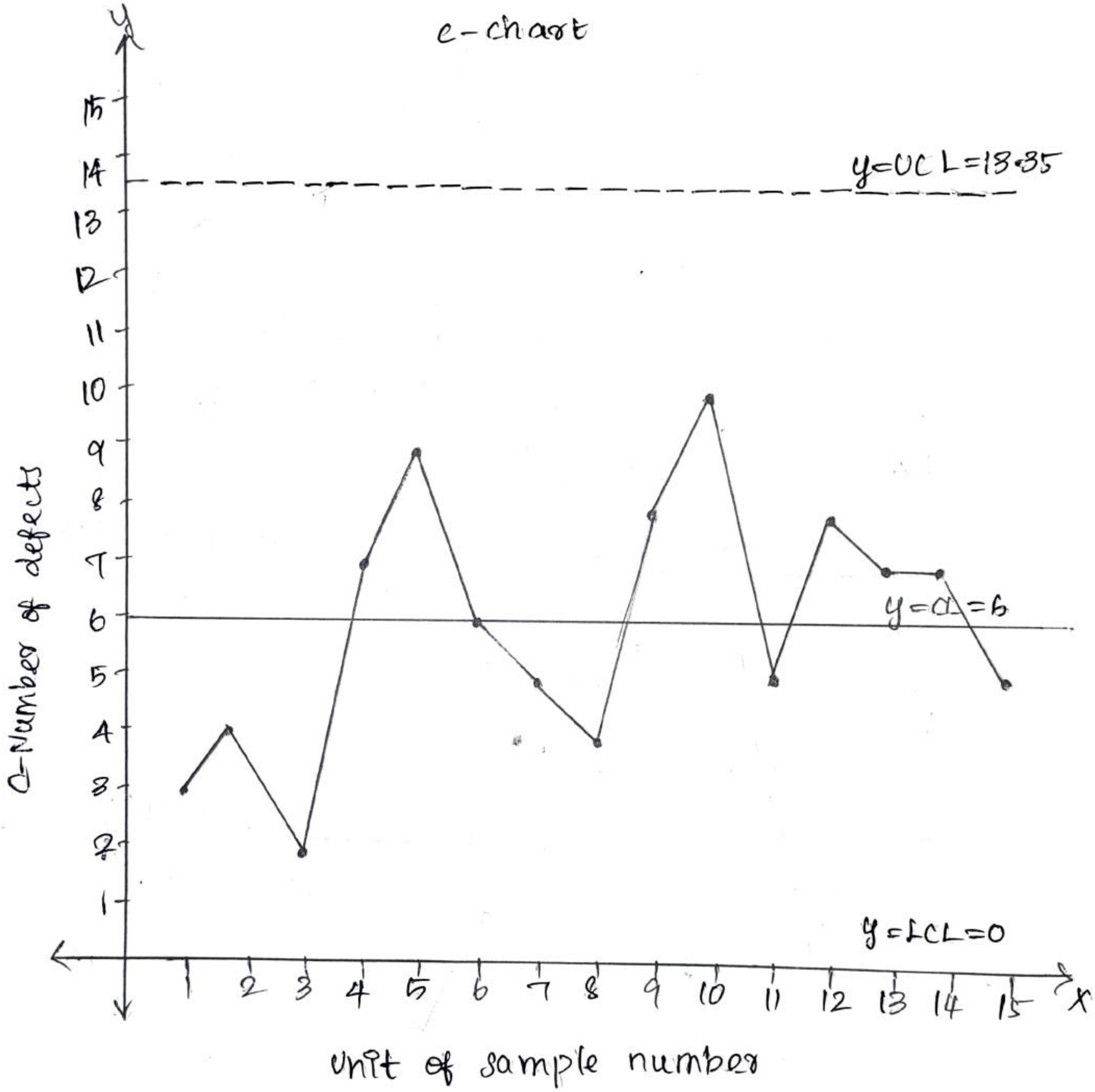
$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 6 + 3\sqrt{6} = 13.35$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 6 - 3\sqrt{6} = -ve.$$

$$\therefore LCL = 0.$$

$$CL = \bar{c} = 6.$$

Scanning the given values of c, we find all the values of c lie between LCL = 0 and UCL = 13.35. Hence the process is under control.



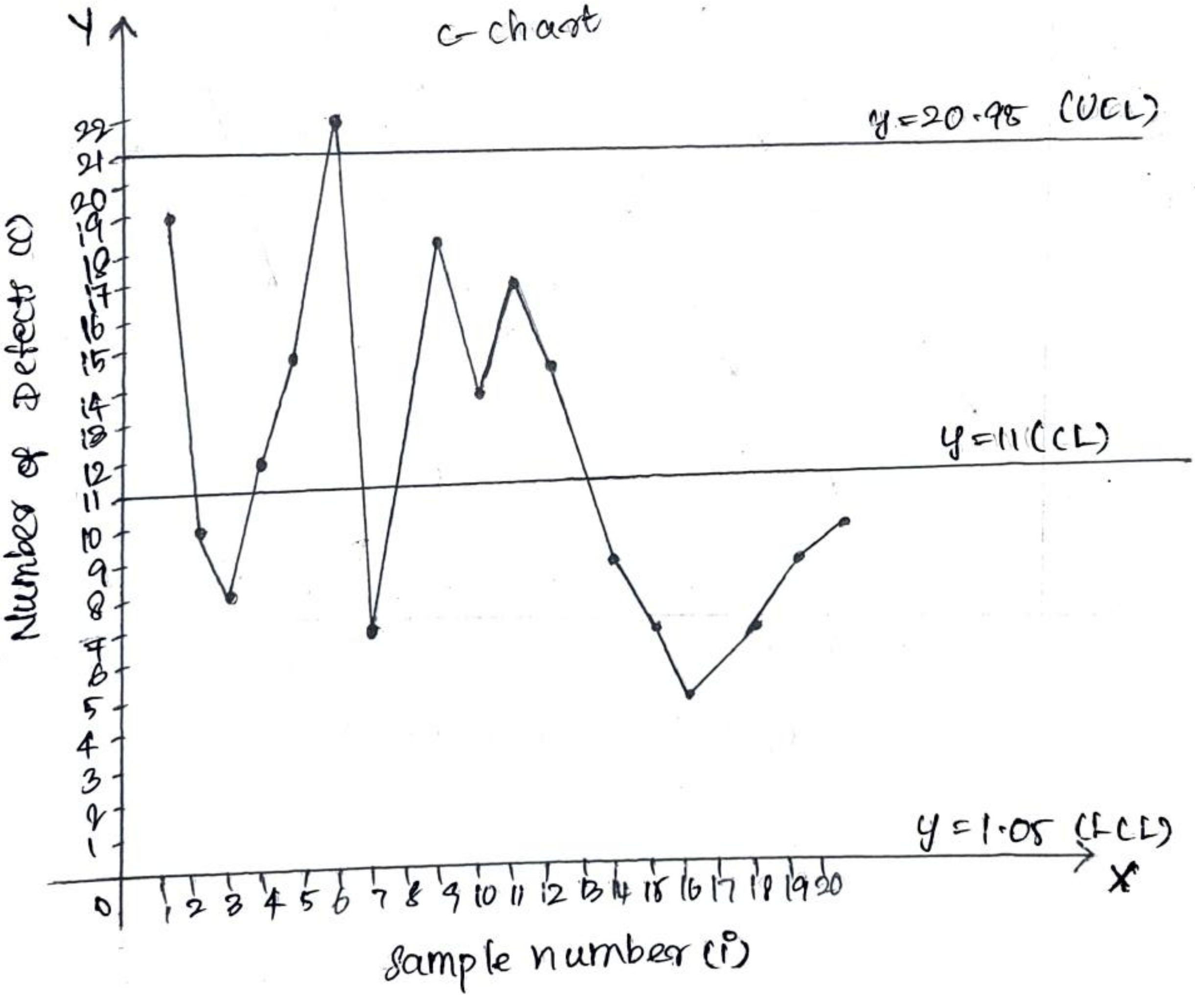
14. A plant produces paper for newsprint and rolls of paper are inspected for defects. The results ~~for~~ of inspection of 20 rolls of paper are given below. Draw c-chart and comment on state of control

Roll no (i) :	1	2	3	4	5	6	7	8	9	10
No. of defects (c) :	19	10	8	12	15	22	7	13	18	13
(i)	11	12	13	14	15	16	17	18	19	20
(c).	16	14	8	7	6	4	5	6	8	9

$\bar{c} = \frac{1}{N} \Sigma c_i = \frac{1}{20} \times 220 = 11$

$CL = \bar{c} = 11;\quad LCL = \bar{c} - 3\sqrt{\bar{c}} = 11 - 3\sqrt{11} = 1.05$

$UCL = \bar{c} + 3\sqrt{\bar{c}} = 11 + 3\sqrt{11} = 20.95$



Since one point falls outside, the process is out of control

15. construct a control chart for defectiveness for the following data.

sample No	1	2	3	4	5	6	7	8	9	10
No. inspected	90	65	85	70	80	80	70	95	90	75
No. of defectives	9	7	3	2	9	5	3	9	6	7

Soln:-

We note that the size of the sample varies from sample to sample. We can construct p-chart provided $0.75\bar{n} < n_i < 1.25\bar{n}$, for all i.

Here, $\bar{n} = \frac{1}{N}\sum n_i = \frac{1}{10}(90+65+\ldots+90+75) = \frac{1}{10}(800)$

$= 80.$

$0.75\bar{n} = 60$ and $1.25\bar{n} = 100$

The values of n_i between 60 and 100. Hence p-chart, can be drawn by the method given below.

Now $\bar{p} = \frac{\text{Total no. of defectives}}{\text{Total no. of items inspected}}$

$= \frac{60}{800} = 0.075$

Hence for the p-chart to be constructed,

$CL = \bar{p} = 0.075$

$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} = 0.075 - 3\sqrt{\frac{0.075 \times 0.925}{80}}$

$= -0.013.$

LCL cannot be negative.

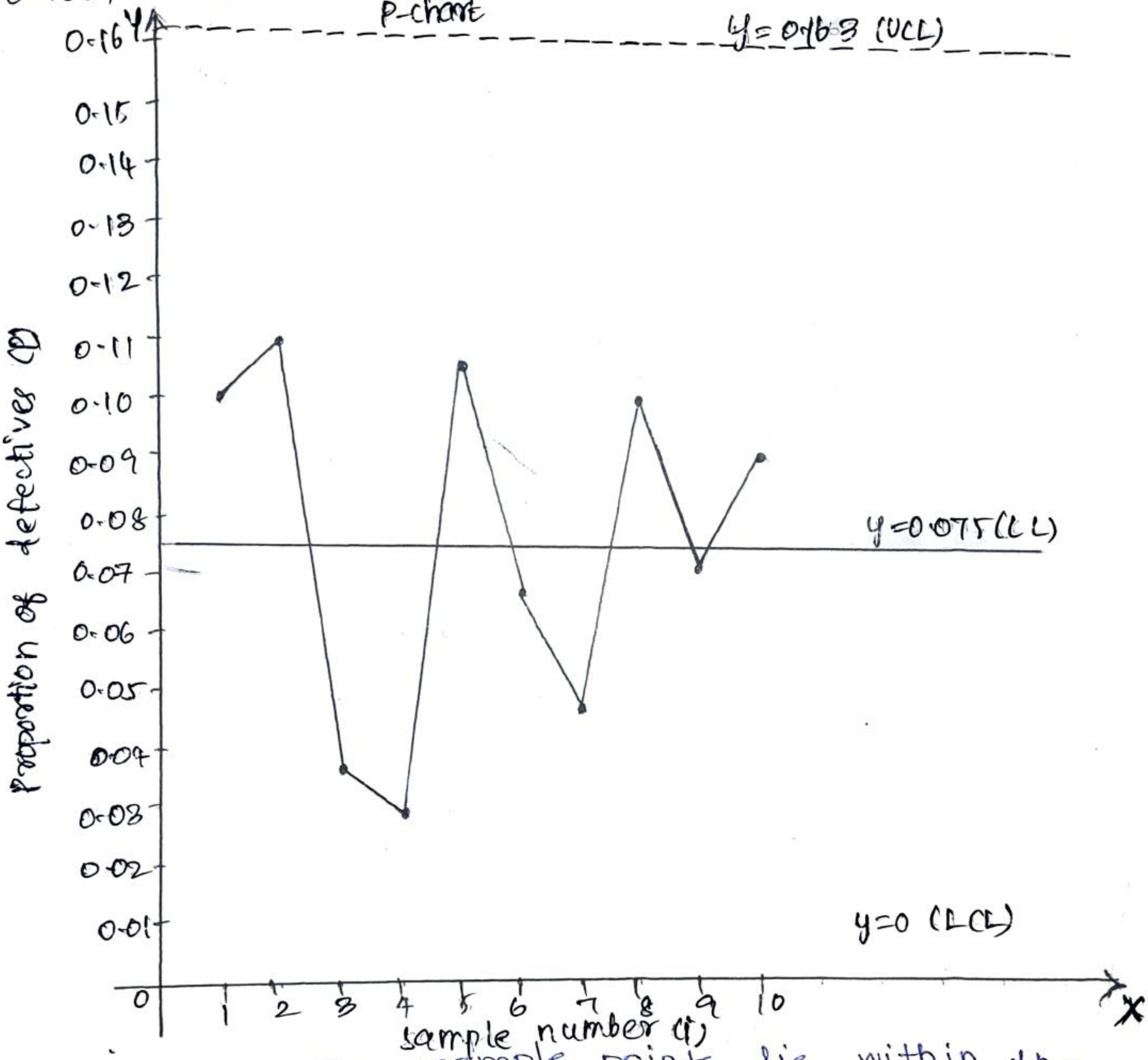
$\therefore LCL = 0$

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{\bar{n}}} = 0.075 + 3\sqrt{\frac{0.075 \times 0.925}{80}}$$

$$= 0.163$$

The values of p_i for the various samples are

0.100, 0.108, 0.035, 0.029, 0.113, 0.043, 0.095, 0.067, 0.093



Since all the sample points lie within the control lines, the process is under control.

16. The following are the figures for the number of defectives of 10 samples each containing 100 items: 8, 10, 9, 8, 10, 11, 7, 9, 6, 12

Draw control chart for fraction defective and comment on the state of control of the process.

soln:- Given the size of all samples are equal.

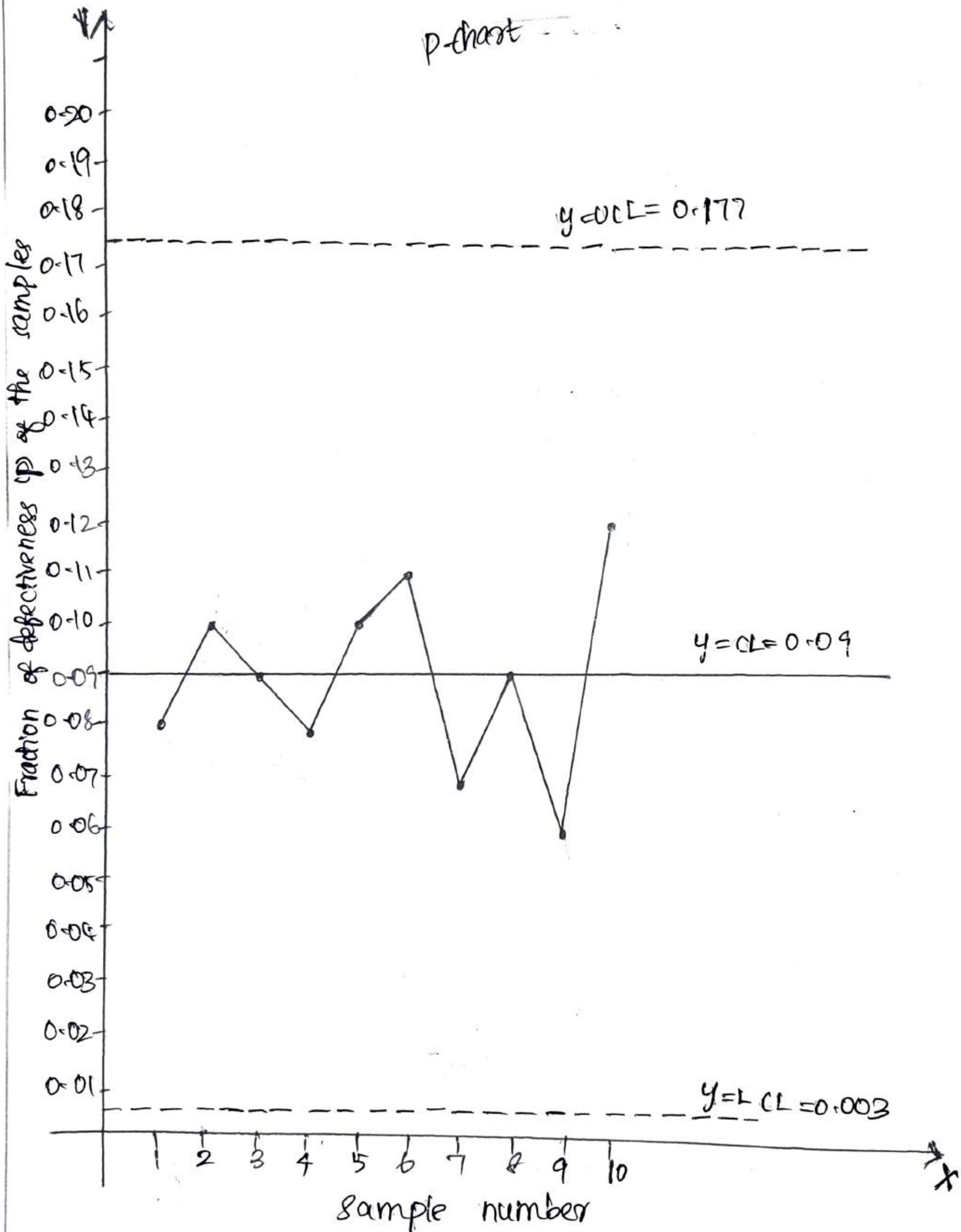
$$p \text{ for sample} = \frac{\text{No. of defectives in sample}}{\text{No. of items in sample}}$$

$$p \text{ for sample } i = \frac{8}{100} = 0.08$$

similarly calculate p for each sample and tabulate. Divide the number of defectives by 100 to get the fraction defective.

sample No.	1	2	3	4	5	6	7	8	9	10
No. of defectives	8	10	9	8	10	11	7	9	6	12
p = fraction defectives:	0.08	0.10	0.09	0.08	0.10	0.11	0.07	0.09	0.06	0.12

p-chart


p-chart
y=UCL= 0.177
y=CL= 0.09
y=LCL =0.003
Fraction of defectiveness (p) of the samples
Sample number
0.20
0.19
0.18
0.17
0.16
0.15
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
1 2 3 4 5 6 7 8 9 10
X
Y

$$\bar{p} = \frac{\Sigma p}{n} = \frac{0.08 + 0.10 + 0.094 \ldots + 0.06 + 0.12}{10}$$

$= 0.09$

(or)

$$\bar{p} = \frac{\text{total no. of defectives in sample}}{\text{total no. of items in sample.}}$$

$$= \frac{90}{10 \times 100} = 0.09$$

since we have 10 samples of size 100

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= 0.09 + 3\sqrt{\frac{0.09 \times 0.91}{100}} = 0.09 + 3(0.029)$$

$= 0.177.$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$= 0.09 - 0.087 = 0.003$

central line corresponds to $\bar{p} = 0.09$

All values of p are > 0.003 and < 0.177

i.e. All sample points lie inside control limits.

The process is under good control.

17. The data given below are the number of defectives in 10 samples of 100 items each. construct a p-chart and np-chart and comment on the results:

Sample No.	1	2	3	4	5	6	7	8	9	10
No. of defectives	6	16	7	3	8	12	7	11	11	4

Soln:-

sample size is constant for all samples, $n = 100$.

total no. of defectives

$= 6+16+7+3+8+12+7+11+11+4 = 85$

total No. Inspected $= 10 \times 100 = 1000$

Average fraction defective $= \bar{p}$

$$\bar{p} = \frac{\text{Total no. of defectives}}{\text{total no. of items inspected}}$$

$$= \frac{85}{1000} = 0.085$$

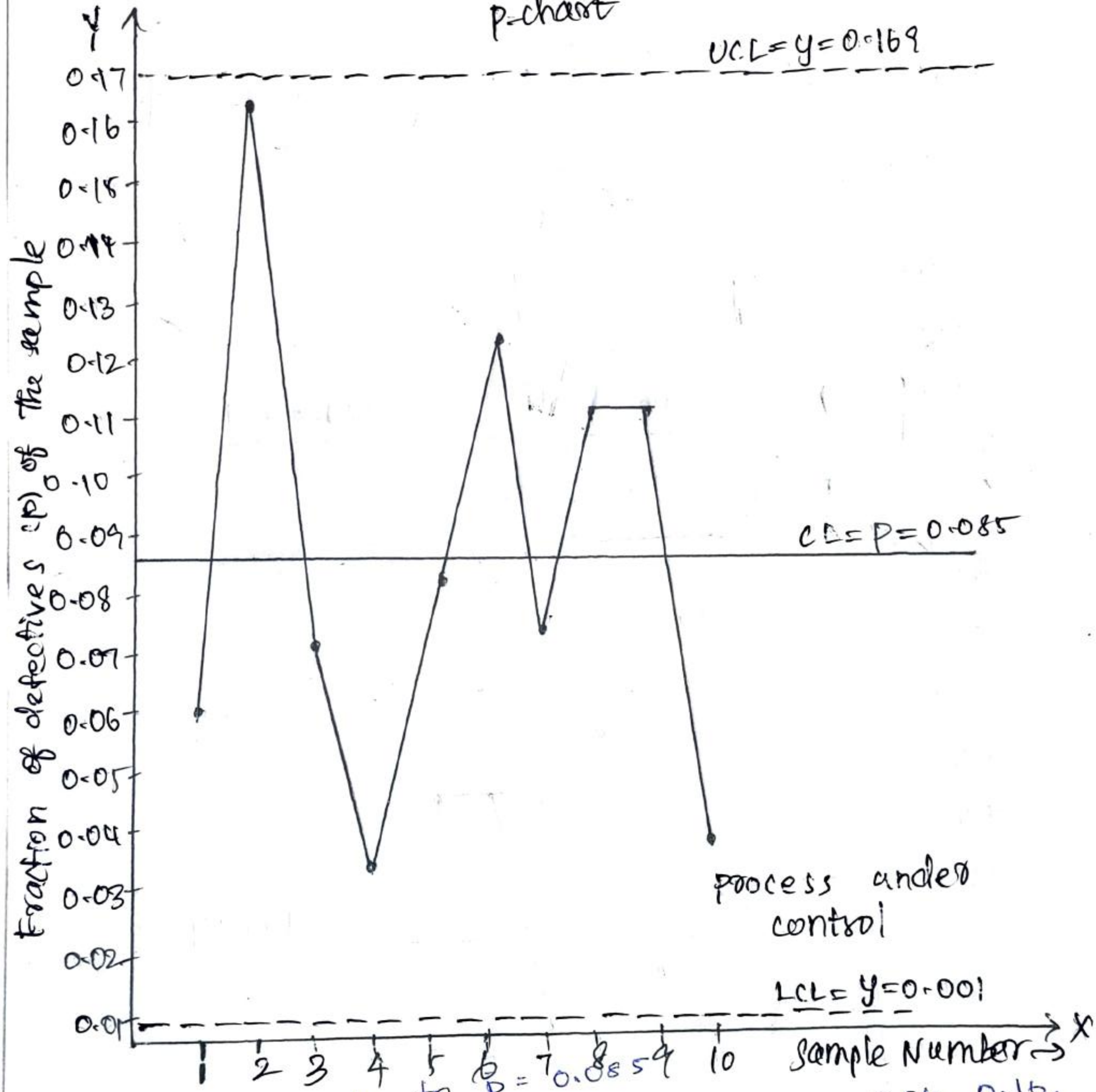
For p-chart:

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 + 3\sqrt{\frac{(0.085)\times(0.915)}{100}}$$

$$= 0.1687$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = 0.085 - 3\sqrt{\frac{0.085 \times 0.915}{100}}$$

$$= 0.0013$$



CL corresponds to $\bar{p}$ = 0.085

Fraction defectives for samples are 0.06, 0.16, 0.07, 0.03, 0.8, 0.12, 0.07, 0.11, 0.11, 0.04

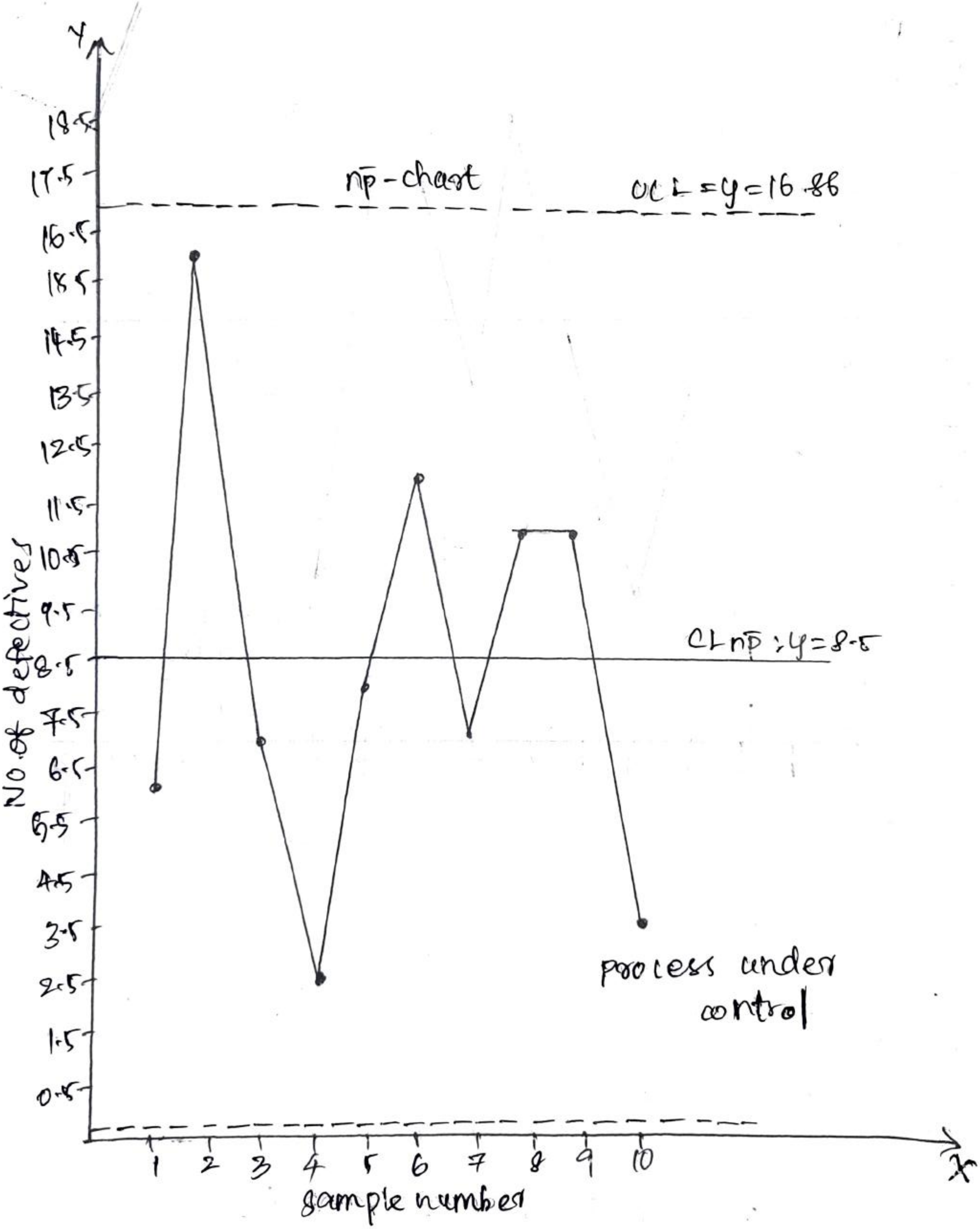
Conclusion: All these values are less than UCL and less than LCL.

The process is under statistical control.

For np-chart:

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} = n\left[\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

$$= 100 \times 0.1687 = 16.87$$



$$n\bar{p} = 100 \times 0.085 = 8.5$$

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

$$= n\left[\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right]$$

$$= 100\,(0.0013) = 0.13$$

Conclusion:

All the values of number of defectives in the table lie between 16.87 and 0.13

Hence the process is under control even in np-chart