SEMESTER : I
ALLIED COURSE : I - Mathematics

Inst Hour	: 5
Credit	: 3
Code	: 18K1CSAM1

PROBABILITY AND STATISTICS

(For B.Sc., Computer Science Major)

UNIT 1:
Theory of Probability – Different definitions of probability sample space – Probability of an event - Independence of events
(Chapter 4: sec4.3, 4.3.1, 4.3.2, 4.5.1, 4.5.2, 4.7.3)

UNIT 2:
Random variables – Distribution functions – Discrete & continuous random variables – Probability mass & density functions
(Chapter 5: Sec 5.1-5.4)

UNIT 3:
Expectation – Variance – Covariance (Chapter 6: Sec 6.1- 6.7)

UNIT 4:
Correlation & Regression – Properties of Correlation & regression coefficients – Angle between two lines of regression - Numerical Problems for finding the correlation & regression coefficients.
(Chapter 10:Sec10.1-10.4, 10.7.2-10.7.5) 11.4, 11.2.1, 11.2.2, 11.2.3

UNIT 5 : 10.2, 10.3, 10.4, 10.5
Theoretical Discrete & Continuous distributions – Binomial, Poisson , Normal distributions- Moment generating functions of these distributions – additive properties of these distributions - and mean for the Binomial, Poisson and Normal distributions (simple problems)
(Chapter 7: Sec7.1,7.2,7.2.1,7.2.4,7.2.6,7.2.7,7.3.0,7.3.2,7.3.4,7.3.5,7.3.8 & Chapter 8- Topics Relevant to normal Distribution)

Text Book :

[1] Gupta.S.C &Kapoor,V.K, Fundamentals of Mathematical Statistics, Sultan Chand & Sons, NewDelhi - 2000 Edition

Reference Book

[1] Thambidurai .P., Practical Statistics , Rainbow publishers – CBE (1991)

Question Pattern

Section A : 10 x 2 = 20 Marks, 2 Questions from each Unit.
Section B : 5 x 5 = 25 Marks, EITHER OR (a or b) Pattern, One question from each Unit.
Section C : 3 x 10 = 30 Marks, 3 out of 5, One Question from each Unit.

9.3.18
HoD
THANJAVUR-613 007

9/3/18

9/3/18

Theory of Probability :

Random Experiment : If in each trial of an experiment conducted under identical conditions, the outcome is not unique, but may be any one of the possible outcomes, then such an experiment is called a random experiment.

Ex: random experiments are tossing a coin, throwing a die, etc.,

Outcome : The result of a random experiment will be called an outcome.

Trial and Event : Any particular performance of a random experiment is called a trial and outcome or combination of outcomes are termed as events.

Ex: Tossing of a coin is a random experiment or trial and getting of a head or tail is an event.

Exhaustive Events : The total number of possible outcomes of a random experiment is known as exhaustive events.

Ex: In tossing of a coin there are two exhaustive events.

Favourable Events: The number of Cases favourable to an event in a trial is the number of outcomes Which entail the happening of the event.

Ex: In throwing of two dice, the number of Cases favourable to getting the Sum 5 is
(1, 4), (4, 1), (2, 3), (3, 2) (ie) 4

Mutually exclusive Events: Events are Said to be mutually exclusive if no two or more of them Can happen Simultaneously in the Same trial.

Ex: In throwing a die all the 6 faces numbered 1 to 6 are mutually exclusive.
In tossing a Coin the events head and tail are mutually exclusive.

Equally Likely Events: Out Comes of trial are Said to be equally likely if taking into Consideration all the relevant evidences, There is no reason to expect one in Preference to the others.

Ex: In throwing an unbiased die, all the Six faces are equally likely to Come.

Independent Events: Several events are said to be independent if the happening of an event is not affected by the supplementary knowledge concerning the occurence of any number of the remaining events.

Ex: In tossing an unbiased coin, the event of getting a head in the first toss is indepenedent of getting a head in the second, third and subsequent throws.

- When a die is thrown twice, the result of the first throw does not affect the result of the second throw.

Probability of an event:

If a random experiment results in 'n' exhaustive mutually exclusive and equally likely events, out of which m are favourable to the occurence of an event E, then the probability 'p' of occurence of E is denoted by P(E) it is defined as

$$P(E) = \frac{\text{Number of favourable events}}{\text{Total number of exhaustive events}} = \frac{m}{n}$$

Note :

(i) $0 \leq P(E) \leq 1$

(ii) $P(E) + P(\bar{E}) = 1$

(iii) Probability 'P' of the happening of an event is known as the probability of success and the probability 'q' of the non-happening of the event as the probability of failure. (ie) $P+q=1$.

(iv) If $P(E)=1$, E is called a Certain event

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

(v) The probability can be computed prior to obtaining any experimental data, it is also called as 'a priori' or mathematical probability.

(vi) The total possible outcomes of a random experiment is called Sample Space

(vii) Each possible outcome in a sample space is called Sample Point.

(vii) The number of sample points in the sample spaces are denoted by $n(S)$.

(viii) Every non-empty subset A of S, which is a disjoint union of single element subsets of the sample space S of a random experiment is called an <u>event</u>

<u>Acceptable assignment of Probabilities</u> :

Let $e_1, e_2 -- e_N$ be mutually disjoint and exhaustive out comes of a random experiment so that its sample space S is $\{e_1, e_2 -- e_N\}$

To each elementary event e_i belonging to S, let us assign a real number called the Probability of the elementary event e_i it is denoted by $P(e_i)$ such that

(i) The Probability of each elementary event is non-negative real number (ie) $P(e_i) \geq 0$ for $i = 1, 2 - N$

(ii) The sum of the Probabilities assigned to all elementary events of the sample space is 1.

(ie) $\sum_{i=1}^{N} P(e_i) = 1$

Such an assignment of real nos. to the elementary events of the sample space is called acceptable assignment of Probabilities.

Probability function: P(A) is the Probability function defined on a σ field B of events if the following Properties or axioms hold.

(i) For each $A \in B$ $P(A)$ is defined, is real and $P(A) \geq 0$

(ii) $P(S) = 1$

(iii) If $\{A_n\}$ is finite or infinite Sequence of disjoint events in B then

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

Independent events: An event A is Said to be independent of another event B, if the Conditional Probability of A given B is equal to the unconditional Probabity of B. (ie) if $P[A/B] = P[A)$

Similarly $P[B/A] = P(B)$; $P(A) \neq 0$

Note: If A and B are independent events then $P[A \cap B] = P(A) \cdot P(B)$

Problems : ① Find the Probability of getting a head (H) in tossing a coin.

② Find the Probability of getting 2 in throwing a die.

③ Find the Probability of getting a tail in tossing a coin.

④ Find the Probability of throwing (i) 4 (ii) an odd number (iii) an even number with an ordinary die.

⑤ Find the Probability of throwing 7 with two dice.

⑥ A bag contains 6 red and 7 black balls. Find the Probability of drawing a red ball.

⑦ Find the Probability that if a card is drawn at random from an ordinary pack, it is a diamond.

⑧ From a pack of 52 cards, one card is drawn at random. Find the Probability of getting a queen.

⑨ Four persons are chosen at random from a group containing 3 men, 2 women and 4 children. S.T the chance that exactly two of them will be children is $10/21$.

Total no. of Persons = 9

Let 4 Persons are choosen at random.

(ie) $n(S) = {}^9C_4 = \frac{9\times8\times7\times6}{1\times2\times3\times4} = 126$ ways

Let A be a favourable event with exactly two of them will be children.

(ie) $n(A) = {}^4C_2 \times {}^5C_2 = 60$

$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{60}{126} = \frac{10}{21}$

(10) From a group of 3 Indians, 4 Pakistanis and 5 Americans, a Sub-Committee of four People is Selected by lots. Find the Probability that the Sub-Committee will Consists of

(i) 2 Indians and 2 Pakistanis

(ii) 1 Indian, 1 Pakistani and 2 Americans

(iii) 4 Americans.

<u>Ans:</u> (i) $n(S) = {}^{12}C_4$; $n(A) = {}^3C_2 \times {}^4C_2$; $P(A) = 18/495$

(ii) $n(S) = {}^{12}C_4$; $n(B) = {}^3C_1 \times {}^4C_1 \times {}^5C_2 = 120$

$P(B) = \frac{120}{495} = \frac{24}{99}$

(iii) $n(S) = {}^{12}C_4$; $n(C) = {}^5C_4$; $P(C) = \frac{5}{495} = \frac{1}{99}$.

(11) A Bag Contains 7 white, 6 red and 5 black balls. Two balls are drawn at random. Find the Probability that they will both be white.

Ans: $n(S) = {}^{18}C_2$; $n(A) = {}^{7}C_2$; $P(A) = \frac{21}{153}$

(12) What is the Probability of having a King and a Queen when two Cards are drawn from a Pack of 52 Cards.

Ans: $n(S) = {}^{52}C_2$; $n(A) = {}^{4}C_1 \cdot {}^{4}C_1$; $P(A) = 8/663$

(13) What is the Probability that of 6 Cards taken from a full Pack, 3 will be black and 3 will be red.

Ans: $n(S) = {}^{52}C_6$; $n(A) = {}^{26}C_3 \cdot {}^{26}C_3$; $P(A) = \frac{n(A)}{n(S)}$

(14) Find the Probability that a hand at bridge will Consist of 3 Spades, 5 hearts, 2 diamonds and 3 clever.

Ans: $n(S) = {}^{52}C_{13}$; $n(A) = {}^{13}C_3 \cdot {}^{13}C_5 \cdot {}^{13}C_2 \cdot {}^{13}C_3$.

(15) What is the Chance that a leap year Selected at random will Contain 53 Sundays?

Ans: $P(A) = 2/7$ $\left[\because \frac{366}{7} \text{ days} = 52 \text{ weeks } 2 \text{ days}\right]$

Theorems on Probability

① Probability of the impossible event is zero

(i.e) $P(\phi) = 0$

Pf/ The certain event 'S' and the impossible event ϕ are mutually exclusive.

(i.e) $S \cup \phi = S$

$P[S \cup \phi] = P(S) \Rightarrow P(S) + P(\phi) = P(S)$

$\Rightarrow P(\phi) = 0$

② Probability of the complementary event $\bar{A}$ of A is given by $P(\bar{A}) = 1 - P(A)$

Pf/. W.K.T A and $\bar{A}$ are disjoint

Since $A \cup \bar{A} = S$

$P(A \cup \bar{A}) = P(S) = 1$

$P(A) + P(\bar{A}) = 1 \Rightarrow P(\bar{A}) = 1 - P(A)$

③ For any two events A and B

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

Pf. We have $\bar{A} \cap B$ and $A \cap B$ are disjoint events.

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

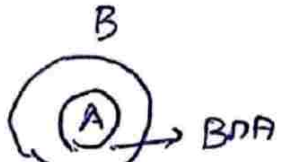
$$P[(A\cap B)\cup(\bar{A}\cap B)] = P(B)$$

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

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

(4) If A and B are two events such that $A \subset B$

P.T $P(B\cap A') = P(B) - P(A)$

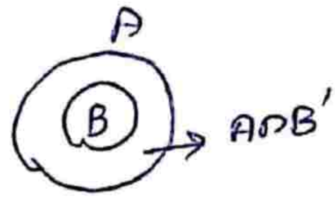


Pf/ Given $A\subset B$

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

$$P(B) = P(A) + P(B\cap A') \quad [\because A \text{ \& } B\cap A' \text{ are mutually exclusive}]$$

(ie) $P(B\cap A') = P(B) - P(A)$



(5) If $B\subset A$ P.T $P(A) > P(B)$

Pf/ Given $B\subset A$

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

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

$$P(A) - P(B) = P(A\cap B') \quad [\because P(A\cap B') \geq 0]$$

(ie) $P(A) - P(B) \geq 0$

$$\therefore P(A) \geq P(B)$$

⑥ If $A \cap B = \phi$ P.T $P(A) \leq P(B')$

Soln. Given: $A \cap B = \phi$

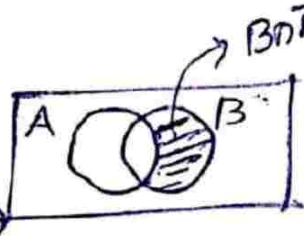
Now $A = (A \cap B) \cup (A \cap B')$

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

$A = A \cap B'$ (i.e) $A \subseteq B' \Rightarrow P(A) \leq P(B')$

⑦ Law of Addition of Probabilities.

Thm: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Where A and B are 2 events and are not disjoint.



Pf: We have $A \cup B = A \cup (B \cap \bar{A})$

[$\because$ A and $B \cap \bar{A}$ are disjoint events]

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

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

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

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

⑧ Multiplication Theorem on Probability

For two events A and B $P(A\cap B) = P(A)\cdot P(B/A)$ $P(A)>0$

$= P(B)\cdot P(A/B)$ $P(B)>0$

Where $P(B/A)$ represents Conditional Probability of occurrence of B when the event A has already happened and $P(A/B)$ is the Conditional Probability of happening A given that B has already happened.

Pf/ W.K.T $P(A) = \frac{n(A)}{n(S)}$; $P(B) = \frac{n(B)}{n(S)}$ & $P(A\cap B) = \frac{n(A\cap B)}{n(S)}$

For the Conditional event A/B the favourable out comes must be one of the sample points of B.

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

$$\Rightarrow P(A/B) = \frac{P(A\cap B)\cdot n(S)}{P(B)\cdot n(S)}$$

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

III'ly $P(A\cap B) = P(A)\cdot P(B/A)$

⑨ If A and B are two events with Positive Probabilities $P(A)\neq 0$ $P(B)\neq 0$ then A & B are indep. $\Leftrightarrow P(A\cap B) = P(A)\cdot P(B)$

Pf/ W.K.T If A and B are indep. Then $P(A/B) = P(A)$ $\rightarrow(1)$; $P(B/A) = P(B)$ $\rightarrow(2)$

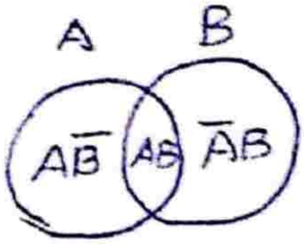
From (1) & (2) $P(A\cap B) = P(A)\cdot P(B)$

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

$\Rightarrow P(B/A) = P(B)$

⑨ If $P(A) = P(B) = P(A \cap B)$ Prove that

$P[A \cap \overline{B} + \overline{A} \cap B] = 0$



Soln/. W.K.T

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

From the fig. $A \cup B = A\overline{B} + AB + \overline{A}B$

$$\therefore P(A \cup B) = P(A\overline{B}) + P(AB) + P(\overline{A}B)$$

$$P(A\overline{B}) + P(\overline{A}B) = P(A \cup B) - P(AB)$$

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

$$= P(AB) + P(AB) - P(AB) - P(AB)$$

$$= 0$$

$$\therefore P(A\overline{B}) + P(\overline{A}B) = 0$$

⑩ If the events A and B are such that $P(A) \neq 0$, $P(B) \neq 0$ and A is independent of B, then B is indep. of A.

Pf/

Since A is indep. of B (ie) $P(A/B) = P(A) \Rightarrow \frac{P(A \cap B)}{P(B)} = P(A)$

$\Rightarrow P(A \cap B) = P(A) \cdot P(B)$

$\Rightarrow \frac{P(B \cap A)}{P(A)} = P(B) \Rightarrow P[B/A] = P(B) \Rightarrow$ B is indep. of A.

Problems :

① A Person is Known to hit the target in 3 out of 4 Shots, Whereas another Person is known to hit the target in 2 out of 3 Shots. Find the Probability of the targets being hit at all when they both Person try.

Soln. The Prob. that the 1st Person hit the target

(ie) $P(A) = 3/4$

The Prob. that the 2nd Person " " "

(ie) $P(B) = 2/3$

The two events are not mutually exclusive

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

$$= P(A) + P(B) - [P(A) \cdot P(B)] = \frac{17}{12} - \frac{6}{12} = \frac{11}{12}$$

② If from a Pack of Cards a Single Card is drawn What is the Probability that either a Spade or King.

Ans: $4/13$. [A and B are not mutually exclusive]

③ If $P(A) = 0.35$, $P(B) = 0.73$, $P(A \cap B) = 0.14$

Find $P(A' \cup B')$

Soln $P(A' \cup B') = 1 - P(A \cap B) = 1 - 0.14 = 0.86$

④ If A and B are independent and $P(A) = 1/3$, $P(B) = 1/4$

Sol/ Since A and B are independent

$$P(A \cap B) = P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$$

⑤ If $P(A) = 0.65$, $P(B) = 0.4$ and $P(A \cap B) = 0.24$ Can A and B are dependent events.

Sol/ W.K.T If A and B are independent then

$$P(A \cap B) = P(A) \cdot P(B) \Rightarrow 0.24 = 0.65 \times 0.4$$

$$\Rightarrow 0.24 \neq .380$$

$\therefore$ A and B are dependent events.

⑥ If A and B are independent events P.T

(i) $\bar{A}$ and B are independent

(ii) A and $\bar{B}$ are independent

(iii) $\bar{A}$ and $\bar{B}$ are independent

Sol/ Since A and B are independent then

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

(i) W.K.T $B = (A \cap B) \cup (\bar{A} \cap B)$ where $A \cap B$ and $\bar{A} \cap B$ are disjoint

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

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

$$= P(B)\left[1 - P(A)\right] = P(B) \cdot P(\bar{A})$$

$\therefore$ $\bar{A}$ and B are independent.

(ii) W.K.T $A = (A \cap B) \cup (A \cap \bar{B})$

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

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

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

$$= P(A)(1 - P(B)) = P(A) \cdot P(\bar{B})$$

$\therefore$ A and $\bar{B}$ are independent.

(iii) W.K.T $\overline{A \cup B} = \bar{A} \cap \bar{B}$

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

$$P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$$

$$= 1 - P(A) - P(B) + P(A \cap B)$$

$$= 1 - P(A) - P(B) + P(A) \cdot P(B)$$

$$= 1 - P(A) - P(B)\left[1 - P(A)\right]$$

$$= \left[1 - P(A)\right]\left[1 - P(B)\right] = P(\bar{A}) \cdot P(\bar{B})$$

$\therefore$ $\bar{A}$ and $\bar{B}$ are independent.

(7) The Probability that machine A will be Performing an usual function in 5 years time is 1/4 while the Probability that machine B will still be operating usefully at the end of the same Period is 1/3. Find the Probability that both machines will be Performing as usual function.

<u>Solⁿ</u> P (machine A operating usefully) $= 1/4$

P (machine B " ") $= 1/3$

P (Both A and B will operate usefully) $= P(A) \cdot P(B)$

$= \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$

(8) A bag contains 8 white and 10 black balls. Two balls are drawn in succession. What is the Probability that first is white and 2nd is black.

<u>Solⁿ</u> Total no. of balls = 18

$P(A) = \frac{8}{18}$; $P(B) = \frac{10}{18}$

$P(A \text{ and } B) = P(A) \cdot P(B) = \frac{8}{18} \cdot \frac{10}{18}$

(10) An article manufactured by a Company Consists of two Parts A and B. In the Process of manufacture of Part A, 9 out of 100 likely to be defective. Similarly 5 out of 100 are likely to be defective in the manufacture of Part B. Calculate the Probability that the assembled article will not be defective.

Soln

P (A will be defective) = $9/100$

P (A will not be defective) $= 1 - \frac{9}{100} = \frac{91}{100}$

P (B will be defective) = $5/100$

P (B will not be defective) $= 1 - 5/100 = \frac{95}{100}$

P [assembled article will not be defective]

= P (A will not be defective) . P(B will not be defe.)

$= \frac{91}{100} \times \frac{95}{100} = 0.86$.

(11) From a bag Containing 4 white and 6 black balls two balls are drawn at random. If the balls are drawn one after the other without replacement, find the Probability that

(i) both balls are white

(ii) both balls are black

(iii) the first ball is white and 2^{nd} is black

(iv) One ball is white and other is black.

Soln.

(i) Total no. of balls = 10

P [1st ball is white] $= \frac{4}{10}$

P [2nd ball is white] $= 3/9$

P [both balls are white] $= \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}$

(ii) P [1st ball is black] $= \frac{6}{10}$

P [2nd ball is black] $= 5/9$

P [both balls are black] $= \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$

(iii) P [1st ball is white] $= \frac{4}{10}$

P [2nd ball is black] $= 6/9$

P [1st white and 2nd black] $= \frac{4}{10} \cdot \frac{6}{9} = \frac{4}{15}$

(iv) P[1st ball is white and 2nd is black]

$= \frac{4}{10} \cdot \frac{6}{9} = \frac{24}{90}$

P[1st ball is black and 2nd ball is white]

$= \frac{6}{10} \times \frac{4}{9} = \frac{24}{90}$

Here both events are mutually exclusive

∴ P[one ball is white and the other is black]

$= \frac{24}{90} + \frac{24}{90} = \frac{8}{15}$

(12) Find the Probability in each of the above four cases, if the balls are drawn one after the other with-replacement.

Soln

(i) P[both balls are white] $= \frac{4}{10} \cdot \frac{4}{10} = \frac{4}{25}$

(ii) P[both balls are black] $= \frac{6}{10} \cdot \frac{6}{10} = \frac{9}{25}$

(iii) P[1st white 2nd black] $= \frac{4}{10} \cdot \frac{6}{10} = \frac{6}{25}$

(iv) P[1st ball is white and 2nd ball is black] $= \frac{4}{10} \cdot \frac{6}{10} = \frac{24}{100}$

P[1st ball is black and 2nd is white] $= \frac{6}{10} \cdot \frac{4}{10} = \frac{24}{100}$,

P[One is white & other is black] $= \frac{24}{100} + \frac{24}{100} = \frac{12}{25}$

(13) Four Cards are drawn without replacement. What is the Probability that they are all Aces?

Solo!

$P(A) = P[\text{getting 1st Ace}] = \frac{4}{52}$

$P(B) = P[\text{getting 2nd Ace}] = \frac{3}{51}$

$P(C) = P[\text{getting 3rd Ace}] = \frac{2}{50}$

$P(D) = P[\text{getting 4th Ace}] = \frac{1}{49}$

$\therefore P[\text{all four Cards are Aces}]$

$= P(A) \cdot P(B) \cdot P(C) \cdot P(D) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49}$.

(14) If two dice are thrown what is the Probability that the Sum is (i) greater than 8 (ii) neither 7 nor 11.

Solo (i) $P[S=9] = 4/36$; $P[S=10] = \frac{3}{36}$; $P[S=11] = \frac{2}{36}$

$P[S=12] = \frac{1}{36}$ $\therefore P(S>8) = 5/18$

(ii) $P[\text{neither 7 nor 11}] = P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$

$= 1 - \{P(A) + P(B)\} = 1 - \frac{1}{6} - \frac{1}{18} = 7/9$

UNIT - II

Random Variable:

A real variable 'X' whose value is determined by the outcome of a random experiment is called a Random Variable.

Example : 1

Consider random experiment of throwing a die. Then 'X' the number of points on the die is a random variable, since 'X' takes the values 1, 2, 3, 4, 5, 6.

Here the random variable 'X' takes the value 1, 2, 3, 4, 5 and 6 each with the probability 1/6.

Note that "twice the number of points on a die" which takes the values 2, 4, 6, 8, 10, 12 is also a random variable.

The "square of number of points on a die" which takes the values 1, 4, 9, 16, 25, 36 is also a random variable.

Example : 2

Consider a random experiment of throwing a coin twice.

We have the following result

HH - 2, HT - 1, TH - 1, TT - 0.

The "number of heads" which takes the value 2, +1, 1, 0 is a random variable.

Some Important Theorems on Random Variables:

Theorem: 1.

If X_1 and X_2 are random variables and K is a constant then KX_1, $X_1 + X_2$, $X_1 X_2$, $K_1X_1 + K_2X_2$, $X_1 - X_2$ are also random variables.

Theorem: 2:

If X is a random variable and $f(\cdot)$ is a continuous function, then $f(X)$ is a random variable

---

Theorem: 3

If X is a random variable and $f(\cdot)$ is an increasing function, then $f(X)$ is a random variable.

Distribution Function of the Random Variable X

The distribution function of a random variable X defined in $(-\infty, \infty)$ is given by

$$F(x) = P(X \leq x)$$

Properties of the Distribution function:

Property 1: $P(a < X \leq b) = F(b) - F(a)$

where $F(x) = P(X \leq x)$

Proof:

The events $a < X \leq b$ and $X \leq a$ are disjoint events and their union is the event $X \leq b$.

$\therefore$ (ie) $P(a < X \leq b) + P(X \leq a) = P(X \leq b)$

(Addition law of probality)

$\therefore P(a < X \leq b) = P(X \leq b) - P(X \leq a)$

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

$[\because F(x) = P(X \leq x)]$

Property 2

$P(a \leq X \leq b) = P(X=a) + F(b) - F(a)$

Proof: $P(a \leq X \leq b) = P\{(X=a) \cup (a < X \leq b)\}$

[$\because$ The event $X=a$ and $a < X \leq b$ are disjoint and their union is the event $a \leq X \leq b$]

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

[Using addition law of Probability]

$= P(X=a) + F(b) - F(a)$

[using property (1)]

Property 3:

$P(a < X < b) = P(a < X \leq b) - P(X=b)$

Proof: $P(a < X < b) = P(a < X \leq b) - P(X=b)$

$= F(b) - F(a) - P(X=b)$

[Using property (1)]

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

$= F(b) - F(a) - P(X=b) + P(X=a)$

Note 1: If $F(x)$ is the distribution function of one dimensional random variable, then

(i) $0 \leq F(x) \leq 1$ (ii) If $x < y$ then $F(x) \leq F(y)$

(iii) $F(-\infty) = 0$, $F(\infty) = 1$.

Discrete Random Variable

If the random variable takes the values only on the set $\{0, 1, 2, 3 \ldots n\}$ is called a Discrete Random Variable.

Example:

* The number of printing mistakes in each page of a book
* the number of telephone calls received by the telephone operator

These are examples of Discrete Random Variable.

Probability Mass Function:

Let x be a one-dimensional discrete random variable which takes the values $x_1, x_2, x_3, \ldots$

Let each possible outcome 'x_i' we can associate a number

$$p_i \; [P(x = x_i)] = P(x_i) = p_i$$

The numbers $P(x_i)$, $i = 1, 2, \ldots$ satisfies the following Conditions

(i) $P(x_i) \geq 0$

(ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

This function 'P' satisfying the above two Conditions is called "the probability mass function" or probability function of the random Variable 'x' and the set $\{x_i, P(x_i)\}$ is called the Probability distribution of the random Variable X.

Example: 1

A random variable 'X' has the following probability function

Values of X :	0	1	2	3	4	5	6	7	8
Probability p(x) :	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) Determine the value of 'a'

(ii) Find $P(X<3)$, $P(X \geq 3)$, $P(0<X<5)$

(iii) Find the distribution of X.

Solution:

(i) We know that if P(x) is the probability mass function, then $\sum_{i=1}^{\infty} P(x_i) = 1$.

Here 'i' varies from 0 to 8

(ie) $a + 3a + 5a + 7a + 9a + 11a + 13a + 15a + 17a = 1$

(ie) $81a = 1 \Rightarrow \boxed{a = 1/81}$

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

$= a + 3a + 5a$

$= \frac{1}{81} + \frac{3}{81} + \frac{5}{81} = \frac{9}{81}$

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

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

$= 1 - \frac{9}{81} = \frac{72}{81}$

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

$= 3a + 5a + 7a + 9a$

$= 24a = 24(1/81) = \frac{24}{81}$

$\boxed{P(0<X<5) = \frac{24}{81}}$

(iv) To find the distribution function $F(x)$.

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

Example : 2

Suppose that the random variable 'X' assumes three values 0, 1 and 2 with probabilities $1/3$, $1/6$ and $1/2$ respectively. obtain the distribution function of X

Solution: Given

x	0	1	2
$P(x)$	$1/3$	$1/6$	$1/2$

x	$F(x) = P(X \le x)$	
0	$1/3$	
1	$1/3 + 1/6 = 1/2$	$\because P(X \le 1) = P(X=0) + P(X=1)$
2	$1/2 + 1/2 = 1$	$\because P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

Continuous Random Variable:

A random variable 'x' which takes all possible values in a given interval is called continuous Random variable.

Examples: Age, height, weight, etc., are continuous random variables

Probability Density Function

The probability density function for the continuous random variable 'x' in the interval (a,b) is given by,

$$f(x) = \begin{cases} 0, & x < a \\ \phi(x), & a \le x \le b \\ 0, & x > b \end{cases}$$

Note:

1. $f(x) \ge 0, \ -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. The probability $P(E)$ is given by

$$P(E) = \int_E f(x)dx$$

Cumulative Distribution Function

If $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$, then $F(x)$ is defined as the cumulative Distribution Function (or) distribution function of the continuous random variable X.

Note:

(i) $F'(x) = f(x) \ge 0$

(ii) $F(-\infty) = 0$

(iii) $F(\infty) = 1$

(iv) $P(a \le x \le b) = \int_a^b f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$

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

Example : 1

(i) Is the function defined as follows a density function? $f(x) = e^{-x}, \; x \geq 0$
$= 0, \; x < 0$

(ii) If so determine the probability that the variate having this density will fall in the interval $(1,2)$

(iii) Also find the cumulative probability function $F(2) = ?$

Solution: In the interval $(1,2)$, e^{-x} is always positive

(ie) $f(x) \geq 0$ in $(1,2)$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx$$

$$= \int_{-\infty}^{0} 0dx + \int_{0}^{\infty} e^{-x}dx$$

$$= \left[-e^{-x}\right]_{0}^{\infty} = -e^{-\infty} + 1 = 1.$$

Hence $f(x)$ satisfies the conditions of the density function.

(ii) $$P(1 \leq x \leq 2) = \int_{1}^{2} f(x)dx$$

$$= \int_{1}^{2} e^{-x}dx = \left[-e^{-x}\right]_{1}^{2} = -e^{-2} + e^{-1}$$

$$= 0.368 - 0.135$$

$$= 0.233$$

(iii) Cumulative probability function

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

$$= \int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)dx$$

$$= \int_{-\infty}^{0} 0dx + \int_{0}^{2} e^{-x}dx = \left[-e^{-x}\right]_{0}^{2} = -e^{-2} + 1$$

$$= 1 - 0.135 = 0.865$$

Example: 2

A continuous random variable 'x' has a probability density function $f(x) = 3x^2$, $0 \le x \le 1$. Find 'a' and 'b' such that

(i) $P(X \le a) = P(X > a)$ and (ii) $P(X > b) = 0.05$

Solution:

Given $P(X \le a) = P(X > a)$

Since total probability is 1, we have

$P(X \le a) = \frac{1}{2}$ and $P(X > a) = \frac{1}{2}$

when $P(X \le a) = \frac{1}{2} \Rightarrow \int_0^a f(x)dx = \frac{1}{2}$

$$\Rightarrow \int_0^a 3x^2 dx = \frac{1}{2}$$

$$\Rightarrow \cancel{3}\left[\frac{x^3}{\cancel{3}}\right]_0^a = \frac{1}{2}$$

$$\Rightarrow a^3 = \frac{1}{2} \Rightarrow \boxed{a = \left(\frac{1}{2}\right)^{\frac{1}{3}}}$$

when $P(X > b) = 0.05$

$$\Rightarrow \int_b^1 f(x)dx = 0.05$$

$$\Rightarrow \int_b^1 3x^2 dx = 0.05$$

$$\Rightarrow \cancel{3}\left[\frac{x^3}{\cancel{3}}\right]_b^1 = 0.05$$

$$\Rightarrow 1 - b^3 = \frac{1}{20} \Rightarrow b^3 = 1 - \frac{1}{20} = \frac{19}{20}$$

$$\Rightarrow \boxed{b = \left(\frac{19}{20}\right)^{\frac{1}{3}}}$$

Example: 3

The diameter of an electric cable, say X, is assumed to be a continuous random variable with probability density function $f(x) = 6x(1-x)$, $0 \le x \le 1$.

(i) check that above is a p.d.f

(ii) Determine a number 'b' such that

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

Solution:

(i) In the interval $0 \le x \le 1$, $f(x)$ is always positive.

(i.e) in $0 \le x \le 1$, $f(x) > 0$

$$\int_0^1 f(x)dx = \int_0^1 6x(1-x)dx$$

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

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

$\therefore$ $f(x)$ is a p.d.f of a random variable X.

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

(i.e) $\int_0^b f(x)dx = \int_b^1 f(x)dx$

$$6\int_0^b (x-x^2)dx = 6\int_b^1 (x-x^2)dx$$

$$6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^b = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_b^1$$

$$\cancel{6}\left[\frac{b^2}{2} - \frac{b^3}{3}\right] = \cancel{6}\left[\left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{b^2}{2} - \frac{b^3}{3}\right)\right]$$

$$\Rightarrow 3b^2 - 2b^3 = 1 - 3b^2 + 2b^3$$

$$\Rightarrow 4b^3 - 6b^2 + 1 = 0$$

$$\Rightarrow (2b-1)(2b^2 - 2b - 1) = 0$$

$$\Rightarrow 2b - 1 = 0 \quad (\text{or}) \quad 2b^2 - 2b - 1 = 0$$

$$\Rightarrow b = \frac{1}{2} \quad (\text{or}) \quad b = \frac{2 \pm \sqrt{4-8}}{4}$$

$$= \frac{1 \pm i}{2}$$

Here $b = \frac{1}{2}$ is the real value and

$b = \frac{1 \pm i}{2}$ is imaginary.

$\therefore b = \frac{1}{2}$ which lies in $(0,1)$

Various Measures of Central Tendency, Dispersion, Skewness and Kurtosis for Continuous Distribution

If $f(x)$ is the p.d.f of a random variable 'x' which is defined in the interval (a,b) then

(i) Arithmetic mean $= \int_a^b x\, f(x)\,dx$

(ii) Harmonic mean $= \int_a^b \frac{1}{x} f(x)\,dx$

(iii) Geometric mean 'G' is given by,

$$\log G = \int_a^b \log x\, f(x)\,dx$$

(iv) Moments about origin

$$\mu_r' = \int_a^b x^r f(x)\,dx$$

(v) Moment about any point A

$$\mu_r' = \int_a^b (x-A)^r f(x)\,dx$$

(vi) Moment about mean

$$\mu_r = \int_a^b (x - \text{mean})^r f(x)\,dx$$

(vii) Mean deviation about the mean is

$$M.D = \int_a^b |x - \text{mean}|\, f(x)\,dx$$

Example : 1 A probability curve $y = f(x)$ has a range from 0 to ∞. If $f(x) = e^{-x}$, find the mean and variance and the third moment about mean.

Solution: **Mean** $= \int_0^\infty x\, f(x)\, dx$ $\left[\text{using mean} = \int_a^b x\, f(x)\, dx\right]$

$$= \int_0^\infty x e^{-x}\, dx = \left[x(-e^{-x}) - (e^{-x})\right]_0^\infty$$

$$= 1 \qquad \left[\because e^{-\infty} = 0\right]$$

Variance $\mu_2 = \int_0^\infty (x - \text{mean})^2 f(x)\, dx$

$$= \int_0^\infty (x-1)^2 e^{-x}\, dx$$

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

$$= 1 - 2 + 2 = 1 \qquad [\text{using integration by parts}]$$

Third moment about mean

$$\mu_3 = \int_0^\infty (x-1)^3 e^{-x}\, dx$$

$\left[\text{Using } \mu_r = \int_a^b (x - \text{mean})^r f(x)\, dx\right]$

$$= \left\{(x-1)^3(-e^{-x}) - 3(x-1)^2(e^{-x}) + 6(x-1)(-e^{-x}) - 6(e^{-x})\right\}_0^\infty$$

[using Bernouli's formula for integration]

$$= -1 + 3 - 6 + 6 = 2$$

Example: 2

The length of time (in minutes) that a certain lady speaks on the telephone is found to be random Phenomenon with a probability function specified by the Probability density function $f(x)$ as

$$f(x) = A e^{-x/5}, \text{ for } x \geq 0$$

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

Find the value of A that makes $f(x)$ a p.d.f

Solution: (a) If $f(x)$ is p.d.f, then

$$\int_0^\infty f(x)dx = 1 \quad \text{(ie)} \int_0^\infty Ae^{-x/5}dx = 1$$

$$A\left[\frac{e^{-x/5}}{-1/5}\right]_0^\infty = 1$$

$$\cancel{A\left[0-\frac{5}{5}\right]=1}$$

$$A\left[0-\left(\frac{e^0}{-1/5}\right)\right] = 1$$

$$5A = 1 \Rightarrow \boxed{A = 1/5}$$

Example: 3

Is the function defined as follows probability density function?

$$f(x) = \begin{cases} 0, & \text{if } x<2 \\ \frac{3+2x}{18}, & \text{if } 2 \le x \le 4 \\ 0, & \text{if } x>4 \end{cases}$$

If so find the $P(2 \le x \le 3)$.

Solution: In the interval $2 \le x \le 4$, $f(x) \ge 0$

Now $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{2} f(x)dx + \int_2^4 f(x)dx + \int_4^\infty f(x)dx$

$$= 0 + \int_2^4 \frac{3+2x}{18}dx + 0$$

$$= \frac{1}{18}\left[\frac{(3+2x)^2}{4}\right]_2^4$$

$$= \frac{1}{72}\left[121-49\right] = 72/72 = 1$$

$\therefore f(x)$ is a p.d.f

Now $P(2 \le x \le 3) = \int_2^3 f(x)dx$

$$= \int_2^3 \frac{3+2x}{18}dx = \frac{1}{18}\left[3x+2\frac{x^2}{2}\right]_2^3$$

$$= \frac{1}{18}\left[18-10\right] = \frac{8}{18}$$

Hence the problem

Example 4

If a random variable 'x' has the p.d.f

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

find the mean and variance of X

Solution:

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

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

$$= \frac{1}{2}\left[\int_{-1}^{1} (x^2+x)\, dx\right] = \frac{1}{2}\left[\frac{x^3}{3}+\frac{x^2}{2}\right]_{-1}^{1}$$

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

$$\boxed{\text{Mean} = \frac{1}{3}}$$

$$\text{Variance } \mu_2 = \int_{-1}^{1} (x-\tfrac{1}{3})\left(\frac{x+1}{2}\right) dx$$

$$\left[\mu_2 = \int_a^b (x-\text{mean})^2 f(x)\,dx\right]$$

$$= \frac{1}{6}\int_{-1}^{1} (3x-1)(x+1)\, dx$$

$$= \frac{1}{6}\int_{-1}^{1} (3x^2+2x-1)\, dx$$

$$= \frac{1}{6}\left[(x^3+x^2-x)_{-1}^{1}\right] = \frac{1}{6}(1-1) = 0$$

$$\boxed{\text{Variance} = 0}$$

Example: 5

For the following density function

$f(x) = a e^{-|x|}$, $-\infty < x < \infty$, find (i) the value of 'a'

(ii) mean and variance

Solution: Given $f(x)$ is p.d.f

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

(ie) $\int_{-\infty}^{\infty} a e^{-|x|} dx = 1$ (ie) $a \cdot 2 \int_0^{\infty} e^{-|x|} dx = 1$

$[e^{-|x|}$ is an even function$]$

$\Rightarrow 2a \int_0^{\infty} e^{-x} dx = 1 \Rightarrow 2a \left[-e^{-x}\right]_0^{\infty} = 1$

$\Rightarrow 2a\,[0+1] = 1 \Rightarrow \boxed{a = 1/2}$

(ii) Mean $= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \cdot \frac{1}{2} e^{-|x|} dx$

$= \frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x|} dx = 0$

$[\because x e^{-|x|}$ is an odd fn$]$

Variance $\mu_2 = \int_{-\infty}^{\infty} (x-0)^2 \, \frac{1}{2} e^{-|x|} dx$

$= \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|} dx$

$= 2 \times \frac{1}{2} \int_0^{\infty} x^2 e^{-x} dx$ $[\because x^2 e^{-|x|}$ is an even function$]$

$= \int_0^{\infty} x^2 e^{-x} dx$

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

Example : 6

Find the value of 'K' and hence find mean and variance of the distribution

$dF = K x^2 e^{-x} dx, \quad 0 < x < \infty$

Solution: Given $\frac{dF}{dx} = K x^2 e^{-x}$

(ie) the p.d.f is given by

$f(x) = K x^2 e^{-x}$

Since $f(x)$ is a p.d.f.,

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

(ie) $\int_0^{\infty} K x^2 e^{-x} dx = 1 \Rightarrow K \int_0^{\infty} x^2 e^{-x} dx = 1$

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

$$K(0+2) = 1 \qquad [\because e^{-\infty} = 0;\ e^0 = 1]$$

$$\therefore \boxed{K = 1/2}$$

<u>Mean</u> $= \int_0^{\infty} x\, f(x)\, dx$

$$= \int_0^{\infty} x \cdot \tfrac{1}{2} x^2 e^{-x} dx$$

$$= \tfrac{1}{2}\left[\int_0^{\infty} x^3 (e^{-x}) dx\right]$$

$$= \tfrac{1}{2}\left[x^3(-e^{-x}) - 3x^2(e^{-x}) + 6x(-e^{-x}) - 6(e^{-x})\right]_0^{\infty}$$

$$\boxed{\text{Mean} = 3}$$

<u>Variance</u> $\mu_2 = \int_0^{\infty} (x-3)^2 f(x)\, dx$

$$= \int_0^{\infty} (x-3)^2\, \tfrac{1}{2} x^2 e^{-x} dx$$

$$= \tfrac{1}{2} \int_0^{\infty} (x^2 - 6x + 9)\, x^2 e^{-x} dx$$

$$= \tfrac{1}{2} \int_0^{\infty} (x^4 - 6x^3 + 9x^2) e^{-x} dx$$

$$= \tfrac{1}{2}\Big[(x^4 - 6x^3 + 9x^2)(-e^{-x}) - (4x^3 - 18x^2 + 18x)(e^{-x}) + (12x^2 - 36x + 18)(-e^{-x}) - (24x - 36)(e^{-x}) + (24)(-e^{-x})\Big]_0^{\infty}$$

[using integration by parts]

$$= \tfrac{1}{2}[18 - 36 + 24] = 6/2 = 3$$

<u>Example: 7</u>

A random variable 'x' has the p.d.f

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

Find (i) $P(X < \frac{1}{2})$ (ii) $P(\frac{1}{4} < X < \frac{1}{2})$ (iii) $P\left(\frac{X > 3/4}{X > 1/2}\right)$

Solution:

(i) $$P(X < \tfrac{1}{2}) = \int_0^{1/2} f(x)\,dx$$

$$= \int_0^{1/2} 2x\,dx = \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_0^{1/2} = (\tfrac{1}{2})^2 - 0$$

$$= \tfrac{1}{4}$$

(ii) $$P(\tfrac{1}{4} < x < \tfrac{1}{2}) = \int_{1/4}^{1/2} f(x)\,dx$$

$$= \int_{1/4}^{1/2} 2x\,dx = \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_{1/4}^{1/2}$$

$$= (\tfrac{1}{2})^2 - (\tfrac{1}{4})^2$$

$$= \tfrac{1}{4} - \tfrac{1}{16}$$

$$= \frac{4-1}{16} = \tfrac{3}{16}$$

(iii) $$P(X > \tfrac{3}{4} \mid X > \tfrac{1}{2}) = \frac{P(X > \frac{3}{4})}{P(X > \frac{1}{2})} \longrightarrow (1)$$

Now $$P(X > \tfrac{3}{4}) = \int_{3/4}^{1} f(x)\,dx = \int_{3/4}^{1} 2x\,dx$$

$$= \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_{3/4}^{1}$$

$$= 1 - (\tfrac{3}{4})^2$$

$$= 1 - \tfrac{9}{16} = \tfrac{7}{16} \rightarrow (2)$$

$$P(X > \tfrac{1}{2}) = \int_{1/2}^{1} f(x)\,dx = \int_{1/2}^{1} 2x\,dx$$

$$= \cancel{2}\left[\frac{x^2}{\cancel{2}}\right]_{1/2}^{1}$$

$$= 1 - \tfrac{1}{4} = \tfrac{3}{4} \rightarrow (3)$$

Substituting (2) & (3) in (1), we get

$$P(X > \tfrac{3}{4} \mid X > \tfrac{1}{2}) = \frac{P(X > \frac{3}{4})}{P(X > \frac{1}{2})} = \frac{7/16}{3/4}$$

$$= \tfrac{7}{12}$$

Example : 8

A continuous random variable 'x' is distributed over the interval [0,1] with p.d.f. ax^2+bx, where a, b are constants. If the arithmetic mean of 'x' is 0.5, find the values of 'a' and 'b'.

Solution: Let $f(x) = ax^2 + bx$

Given $f(x)$ is a p.d.f in [0,1]

(ie) $\int_0^1 f(x)dx = 1$

$$\Rightarrow \int_0^1 (ax^2+bx)dx = 1$$

$$\left[a\frac{x^3}{3} + b\frac{x^2}{2}\right]_0^1 = 1$$

$$a\left(\tfrac{1}{3}\right) + b\left(\tfrac{1}{2}\right) = 1 \Rightarrow 2a+3b=6 \rightarrow (1)$$

Now mean $= \int_0^1 x f(x) dx$

$\left[\text{using mean} = \int_a^b x f(x) dx\right]$

$$= \int_0^1 x(ax^2+bx)dx$$

$$= \int_0^1 (ax^3+bx^2)dx = \left[\frac{ax^4}{4} + \frac{bx^3}{3}\right]_0^1$$

$$= \frac{a}{4} + \frac{b}{3}$$

Given mean $= 0.5 = \tfrac{1}{2}$

$$\therefore \tfrac{1}{2} = \frac{a}{4} + \frac{b}{3}$$

$$\tfrac{1}{2} = \frac{3a+4b}{12} \Rightarrow 3a+4b=6 \rightarrow (2)$$

$$2a+3b=6 \rightarrow (3)$$

Solving (2) & (3), we get

$a = -6,\ b = 6$

Example : 9

For the probability density function

$$f(x) = \frac{2(b+x)}{b(a+b)}, \quad -b \le x < 0$$

$$= \frac{2(a-x)}{a(a+b)}, \quad 0 \le x \le a$$

Find the mean

Solution: $\text{Mean} = \int_{-b}^{a} x f(x)\,dx$

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

$$= \frac{2}{b(a+b)} \int_{-b}^{0} (bx+x^2)\,dx + \frac{2}{a(a+b)} \int_{0}^{a} (ax-x^2)\,dx$$

$$= \frac{2}{b(a+b)} \left[\frac{bx^2}{2} + \frac{x^3}{3}\right]_{-b}^{0} + \frac{2}{a(a+b)} \left[\frac{ax^2}{2} - \frac{x^3}{3}\right]_{0}^{a}$$

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

$$= \frac{2}{b(a+b)} \left[\frac{-3b^3+2b^3}{6}\right] + \frac{2}{a(a+b)} \left[\frac{3a^3-2a^3}{6}\right]$$

$$= \frac{\cancel{2}^{1}}{a+b}\left(\frac{-b^2}{\cancel{6}_{3}}\right) + \frac{\cancel{2}^{1}}{a+b}\left(\frac{a^2}{\cancel{6}_{3}}\right)$$

$$= \frac{1}{3(a+b)}(a^2-b^2) = \frac{1}{3(\cancel{a+b})}(a-b)(\cancel{a+b})$$

$$\boxed{\text{Mean} = \frac{a-b}{3}}$$

Example: 10

Prove that the geometric mean G of the distribution $dF = 6(2-x)(x-1)\,dx$, $1 \le x \le 2$ is given by $6\log(16G) = 19$.

Solution: Given $dF = 6(2-x)(x-1)\,dx$

$\therefore$ P d f $f(x) = 6(2-x)(x-1)$

$\log G = \int_1^2 \log x \cdot f(x)dx$

$\left[\text{Using } \log G = \int_a^b \log x \cdot f(x)\,dx\right]$

$= 6\int_1^2 \log x\,(2-x)(x-1)\,dx$

$= -6\int_1^2 (x^2-3x+2)\log x\,dx$

$= -6\left[\int_1^2 \log x\, d\left(\frac{x^3}{3}-\frac{3x^2}{2}+2x\right)\right]$

$= -6\left[\left\{\log x\left(\frac{x^3}{3}-\frac{3x^2}{2}+2x\right)\right\}_1^2 - \int_1^2\left(\frac{x^3}{3}-\frac{3x^2}{2}+2x\right)\frac{1}{x}dx\right]$

$= -6\left[\log 2\left(\frac{8}{3}-3\times 2+4\right) - \int_1^2\left(\frac{x^3}{3}-\frac{3x^2}{2}+2x\right)\frac{1}{x}dx\right]$

$= -6\left[\log 2\left(\frac{8}{3}-3\times 2+4\right) - \int_1^2\left(\frac{x^2}{3}-\frac{3x}{2}+2\right)dx\right]$

$[\because \log(1)=0]$

$= -6\left[\log 2\times\frac{2}{3} - \left(\frac{x^3}{9}-\frac{3x^2}{4}+2x\right)_1^2\right]$

$= -6\left[\frac{2}{3}\log 2 - \left(\frac{8}{9}-3+4-\frac{1}{9}+\frac{3}{4}-2\right)\right]$

$= -4\log 2 + 6\left(\frac{8}{3}-1-\frac{1}{3}+\frac{3}{4}\right)$

$= -4\log 2 + 6\left(\frac{19}{36}\right)$

$\log G = -4\log 2 + \frac{19}{6}$

(i.e) $\log G + 4\log 2 = \frac{19}{6}$ $\quad [\because \log m^n = n\log m]$

$\Rightarrow \log G + \log 2^4 = \frac{19}{6}$ $\quad [\because \log mn = \log m + \log n]$

$\Rightarrow \log 16G = \frac{19}{6}$

(i.e) $\boxed{6\log(16G) = 19}$ Hence the problem.

Continuous Distribution Function

If $f(x)$ is a p.d.f of a random variable 'X' then the function $F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$, $-\infty < x < \infty$

is called the distribution function or cumulative distribution function of the random variable 'X'.

Note: (i) $0 \le F(x) \le 1$, $-\infty < x < \infty$

(ii) $F(-\infty) = 0$, $F(+\infty) = 1$

(iii) $$P(a \le X \le b) = \int_a^b f(x)\,dx$$

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

$$= P(X \le b) - P(X \le a)$$

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

(iv) $F'(x) = \dfrac{dF(x)}{dx} = f(x) \ge 0$ [$f(x)$ is p.d.f]

(or) $dF(x) = f(x)\,dx$

$dF(x)$ is called the probability differential of the random variable X.

Hence by differentiating $F(x)$, we get p.d.f $f(x)$.

$$\therefore F(x) = \int_{-\infty}^{x} f(x)\,dx, \quad 0 < x < b.$$

Example: 1

Verify that the following is a distribution function

$$F(x) = \begin{cases} 0, & x < -a \\ \frac{1}{2}\left(\frac{x}{a}+1\right), & -a \le x \le a \\ 1, & x > 1 \end{cases}$$

Solution: clearly (i) $0 \le F(x) \le 1$

(ii) $F(-\infty) = 0$ [$\because F(x) = 0$, $x < -a$ $\therefore F(-\infty) = 0$]

(iii) $F(\infty)=1$ $\left[\because F(x)=1,\ x>1\right.$

$\left.\therefore F(\infty)=1,\ x>1\right]$

Hence $F(x)$ satisfies all the conditions of a distribution function.

(iv) $f(x)=\dfrac{dF(x)}{dx}=\begin{cases}\frac{1}{2a}, & -a\le x\le a\\ 0, & \text{otherwise}\end{cases}$

Also $\int_{-a}^{a} f(x)dx=\int_{-a}^{a}\frac{1}{2a}dx=\frac{1}{2a}\left[x\right]_{-a}^{a}=\frac{1}{2a}(2a)=1.$

$\therefore f(x)$ is a p.d.f and $F(x)$ is a distribution function.

Example: 2

Suppose that the amount of money that a person has saved is found to be a random variable with

$$F(x)=\begin{cases}\frac{1}{2}e^{-(x/50)^2}, & x<0\\ 1-\frac{1}{2}e^{-(x/50)^2}, & x\ge 0\end{cases}$$

(i) Is $F(\cdot)$ continuous? What is the p.d.f?

(ii) What is the probability that the amount of saving possessed by him will be (i) more than 50, (ii) equal to 50 rupees.

(iii) What is the conditional probability that the amount of savings will be less than Rs. 100 given that it is more than 50?

Solution: The given distribution function is continuous since the value $F(x)$ at $x=0$ is the same.

Probability density function

$$f(x)=\frac{d}{dx}F(x)$$

(i) $$f(x)=\begin{cases}\frac{1}{2}e^{-(x/50)^2}(-2)\left(\frac{x}{50}\right)\left(\frac{1}{50}\right)\\ -\frac{1}{2}e^{-(x/50)^2}\cdot\left(\frac{-2x}{50}\right)\left(\frac{1}{50}\right)\end{cases}$$

$$f(x) = \begin{cases} \frac{-x}{2500} e^{-\left(\frac{x}{50}\right)^2}, & x < 0 \\ \frac{x}{2500} e^{-\left(\frac{x}{50}\right)^2}, & x \geq 0 \end{cases}$$

(ii) Let 'X' be the random variable which represent the amount of savings

$$P(X > 50) = 1 - P[X \leq 50]$$
$$= 1 - F(50)$$
$$= 1 - \left\{1 - \tfrac{1}{2} e^{-1}\right\}$$
$$= \frac{1}{2e} = \frac{1}{2.718} = 0.1839$$

$$P(X = 50) = 0$$

[$\because$ The probability that a continuous variable takes a fixed value is zero]

(iii) Let $P(A) = P(X < 100)$
$$= P(X \leq 100) - P(X = 100)$$
$$= F(100) \qquad [\because P(X = 100) = 0]$$
$$= 1 - \tfrac{1}{2} e^{-4} = 1 - \frac{1}{109.19} = 0.99$$

$$P(B) = P(X > 50) = 0.3679 \text{ [From ii]}$$

$$P(A \cap B) = P\{X \leq 100 \cap X > 50\}$$
$$= P[50 < X < 100]$$
$$= F(100) - F(50) = 0.99 - 0.817$$
$$= 0.173$$

Example : 3

The probability distribution function of a random variable 'X' is $f(x) = \begin{cases} x, & 0 < x \leq 1 \\ 2 - x, & 1 \leq x < 2 \\ 0, & x \geq 2 \end{cases}$

Find the cumulative distribution function of X.

Solution: We know that cumulative distribution function $F(x) = \int_{-\infty}^{x} f(x)\,dx$, when x lies in $0 < x < 1$.

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

$$= \int_{-\infty}^{0} f(x)dx + \int_{0}^{x} f(x)dx$$

$$= 0 + \int_{0}^{x} x\,dx = \left[\frac{x^2}{2}\right]_0^x = \frac{x^2}{2}$$

when x lies in $1 < x \leq 2$

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

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

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

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

$$= 2x - \frac{x^2}{2} - 1$$

when 'x' lies in $x \geq 2$,

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

$$= \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx$$

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

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

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

$$\therefore F(x) = \begin{cases} \frac{x^2}{2}, & \text{in } 0 < x < 1 \\ 2x - \frac{x^2}{2} - 1, & \text{in } 1 < x \leq 2 \\ 1, & \text{in } x > 2 \end{cases}$$

Example: 4

A random variable 'X' has the density function

$$f(x) = K \cdot \frac{1}{1+x^2} \quad \text{in } -\infty < x < \infty$$

$$= 0, \quad \text{otherwise}$$

Find 'K' and the distribution function $F(x)$. Find $P(X \geq 0)$.

Solution:

Since $f(x)$ is p.d.f., we have

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

$$K \int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = 1 \Rightarrow K\left(\tan^{-1} x\right)_{-\infty}^{\infty} = 1$$

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

$$\Rightarrow K\left[\pi/2 + \pi/2\right] = 1$$

$$\Rightarrow K \cdot \pi = 1 \qquad \left[\because \tan^{-1}(\infty) = \frac{\pi}{2}\right.$$

$$\Rightarrow \boxed{K = 1/\pi} \qquad \tan^{-1}(-\infty)$$

$$= -\tan^{-1}(\infty)$$

$$\left. = -\pi/2 \right]$$

$\therefore$ $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$, in $-\infty < x < \infty$

$= 0$, otherwise

To find $F(x)$.

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

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

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

To find $P(X \geq 0)$

$$P(X \geq 0) = 1 - P(X \leq 0) = 1 - F(0)$$

$$= 1 - \frac{1}{\pi}\left[0 + \pi/2\right] = 1/2$$

Example: 5

The length (in hours) X of a certain type of light bulb may be supposed to be a Continuous random variable with p.d.f

$$f(x) = \frac{a}{x^3}, \quad 1500 < x < 2500$$

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

Determine the constant 'a' the distribution function of X, and compute the probability of the event $1{,}700 \leq X \leq 1900$.

Solution:

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

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

(ie) $$a\int_{1500}^{2500} \frac{1}{x^3}\,dx = 1$$

$$\Rightarrow a\left[-\frac{1}{2x^2}\right]_{1500}^{2500} = 1$$

$$\Rightarrow a\left[\frac{1}{-2(2500)^2} + \frac{1}{2(1500)^2}\right] = 1$$

$$\Rightarrow \boxed{a = 70{,}31{,}250}$$

To find the d.f $F(x)$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-\infty}^{1500} f(x)\,dx + \int_{1500}^{x} f(x)\,dx$$

$$= 0 + \int_{1500}^{x} \frac{a}{x^3}\,dx = \left[-\frac{1}{2x^2}\right]_{1500}^{x}$$

$$= a\left[-\frac{1}{2x^2} + \frac{1}{2(1500)^2}\right] = \frac{a}{2}\left[\frac{1}{(1500)^2} - \frac{1}{x^2}\right]$$

To find $P(1700 \le X \le 1900)$

$$P(1700 \le X \le 1900) = F(1900) - F(1700)$$

$$= \frac{a}{2}\left[\frac{1}{2890000} - \frac{1}{3610000}\right]$$

Hence the problem.

UNIT - II

Mathematical Expectations

Let X be a continuous random variable with probability density function $f(x)$.

Then the mathematical expectation of 'X' is denoted by $E(X)$ and is given by

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\,dx \quad \text{[for continuous random variable]}$$

$$= \sum_{x} x f(x) \quad \text{[for discrete random variable]}$$

r^{th} moment (about Origin)

For the probability distribution, $f(x)$, the r^{th} moment (about origin) is defined as $\mu_r' = \int_{-\infty}^{\infty} x^r f(x)\,dx$

$$\boxed{\mu_r' = E(X^r)}$$

Thus $\mu_1' = E(X)$

$\mu_2' = E(X^2)$

$\therefore$ Mean $= \bar{x} = \mu_1' = E(X)$

and Variance $= \mu_2 = \mu_2' - \mu_1'^2$

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

r^{th} moment about mean

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

$$\boxed{\mu_r = \int_{-\infty}^{\infty} \{x - \bar{x}\}^r f(x)\,dx}$$

This gives the r^{th} moment about mean and it is denoted by μ_r

Thus $\mu_r = \int_{-\infty}^{\infty} (x - \bar{x})^r f(x)\,dx$

Put $r = 1$, we get $\mu_1 = \int_{-\infty}^{\infty} (x - \bar{x}) f(x)\,dx$

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

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

$$= \bar{x} - \bar{x} \quad \left[\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right]$$

$$\boxed{\mu_1 = 0}$$

Put $r = 2$, we get Variance $= \mu_2 = E\left[\{X - E(X)\}^2\right]$

$$\boxed{\mu_2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x)}$$

Problem: Let $g(x) = K$ (constant) then find $E[g(x)]$

Solution: $E[g(x)] = E(K) = \int_{-\infty}^{\infty} k\, f(x)\,dx$

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

$$= K \cdot 1 = K$$

$$\boxed{E(K) = K}$$

For discrete random variables 'X'

$$E(X^r) = \sum_x x^r f(x)$$

(ie) $\mu_r' = E(X^r) = \sum_x x^r f(x)$.

Put $r=1$, we get Mean $= \mu_1' = \sum x\, f(x)$

Variance $= \mu_2 = \mu_2' - \mu_1'^2$

$= E(X^2) - \{E(X)\}^2$

The r^{th} moment about mean

$$\mu_r = E\left[\{X - E(X)\}^r\right]$$

$$= \sum_x (x - \bar{x})^r \cdot f(x) \qquad [\because E(X) = \bar{x}]$$

Put $r=2$, we get

Variance $= \mu_2 = \sum (x - \bar{x})^2 \cdot f(x)$.

<u>Addition Theorem</u> (Expectation)

If X and Y are random Variables, then

$E(X+Y) = E(X) + E(Y)$

<u>Proof</u>: Let X and Y be continuous random Variables with marginal p.d.f $f_X(x)$ and $f_Y(y)$ and whose joint p.d.f $f_{XY}(x,y)$.

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

$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy$

Now $E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f_{XY}(x,y)\,dx\,dy$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f_{XY}(x,y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f_{XY}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f_{XY}(x,y)dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f_{XY}(x,y)dx \right] dy$$

$$E(X+Y) = \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy$$

$$= E(X) + E(Y)$$

[Using mass density function of Y, $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$

mass density function of X, $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$]

Hence the proof.

Multiplication Theorem of Expectations:

If X and Y are independent variables, then

$E(XY) = E(X) \cdot E(Y)$.

Proof: Let X and Y be continuous random variables with joint p.d.f $f_{XY}(x,y)$ and marginal p.d.f's $f_X(x)$ and $f_Y(y)$ respectively.

We know that $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$$

$$\text{Now } E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x,y) dx dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) \cdot f_Y(y) dx dy$$

[Since X and Y are independent]

$$= \int_{-\infty}^{\infty} x f_X(x) dx \cdot \int_{-\infty}^{\infty} y f_Y(y) dy$$

$$\boxed{E(XY) = E(X) \cdot E(Y)}$$

Note: If $X_1, X_2, \dots X_n$ are 'n' independent random variables, then $E[X_1, X_2 \dots X_n] = E(X_1) \cdot E(X_2) \dots E(X_n)$.

Theorem: If 'X' is a random variable and 'a' is a constant, then (i) $E[aG(X)] = aE[G(X)]$

(ii) $E[G(X) + a] = E[G(X)] + a$.

where $G(X)$ is a function of 'X' which is also a random variable.

Proof (i) $E[a\,G(x)] = \int_{-\infty}^{\infty} a\,G(x)\,f(x)\,dx$

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

$$= a\,E[G(x)]$$

(ii) $E[G(x)+a] = \int_{-\infty}^{\infty} [G(x)+a]\,f(x)\,dx$

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

$$= E[G(x)] + a \qquad \left[\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right]$$

Theorem: If X is a random variable and 'a' and 'b' are constants then $E[aX+b] = aE(X)+b$.

Proof: $E[aX+b] = \int_{-\infty}^{\infty} (ax+b)\,f(x)\,dx$

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

$$\boxed{E[aX+b] = a\,E(X)+b}$$

Corollary 1: If $b=0$, then we get $E[aX] = a\,E[X]$

Corollary 2: If we take $a=1$ and $b = -E(X) = -\bar{X}$, then we get $E[X-\bar{X}] = E(X) - E(X) = 0$

Corollary 3: Let $g(X) = aX+b$

$$\therefore E[g(X)] = E[aX+b] = a\,E(X)+b \longrightarrow (1)$$

But $g[E(X)] = a\,E(X)+b \longrightarrow (2)$

From (1) & (2), we get $E[g(X)] = g\,E[X]$

Note: $E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}$; $E[\log(X)] \neq \log E(X)$

$$E(X^2) \neq [E(X)]^2$$

Expectation of a Linear Combination of Random Variables

Let $X_1, X_2, \ldots X_n$ be any 'n' random variables and if $C_1, C_2, \ldots C_n$ are constants, then

$$E[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

Theorem: Let X and Y be two random Variables such that $Y \leq X$, then $E(Y) \leq E(X)$

Proof: Given $Y \leq X \Rightarrow 0 \leq X - Y \Rightarrow X - Y \geq 0$

$$E(X-Y) \geq E(0)$$

$$E(X) - E(Y) \geq 0$$

$$E(X) \geq E(Y)$$

$$\boxed{E(Y) \leq E(X)}$$

Note: $|E(X)| \leq E|X|$

Show that if X is a random Variable, then $V(aX+b) = a^2 V(X)$ where 'a' and 'b' are Constants.

Proof: Let $Y = aX + b$ Then $E(Y) = aE(X) + b$

$$Y - E(Y) = a[X - E(X)] \Rightarrow \{Y - E(Y)\}^2 = a^2\{X - E(X)\}^2$$

$$E\{Y - E(Y)\}^2 = a^2 E\{X - E(X)\}^2$$

$$\Rightarrow V(Y) = a^2 V(X) \quad \text{using } V(X) = E[X - E(X)]^2$$

$$\Rightarrow V(aX+b) = a^2 V(X) \quad [\because Y = aX + b]$$

Hence $\boxed{V(aX+b) = a^2 V(X)}$

where 'a' and 'b' are constants.

Definition: Covariance:

If X and Y are random variables, then covariance between them is defined as

$$\text{cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$$

$$= E\{XY - XE(Y) - E(X)Y + E(X)E(Y)\}$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$\boxed{\text{cov}(X,Y) = E(XY) - E(X) \cdot E(Y)}$$

Remarks: If X and Y are independent, then

$$E(XY) = E(X)E(Y)$$

$$\text{cov}(X,Y) = E(XY) - E(X)E(Y)$$

$$= E(X)E(Y) - E(X)E(Y)$$

$$\boxed{\text{cov}(X,Y) = 0}$$

$\therefore$ If X and Y are independent, then $\text{Cov}(X,Y) = 0$

Note: 1. $\text{cov}(aX, bY) = ab\,\text{cov}(X,Y)$

2. $\text{cov}(X+a, Y+b) = \text{cov}(X,Y)$

3. $\text{cov}(aX+b, cY+d) = ac\,\text{cov}(X,Y)$

4. $V(X_1+X_2) = V(X_1)+V(X_2)+2\,\text{cov}(X_1,X_2)$

5. $V(X_1-X_2) = V(X_1)+V(X_2)-2\,\text{cov}(X_1,X_2)$

If X_1 and X_2 are independent

$$V(X_1 \pm X_2) = V(X_1)+V(X_2) \pm 2\,\text{cov}(X_1,X_2)$$

Moment Generating Function:

The moment generating function of a random variable 'x' whose probability function $f(x)$ is given by

$$M_X(t) = E(e^{tX})$$

$$= \begin{cases} \int_{-\infty}^{\infty} e^{tx} f(x)\,dx, & \text{for continuous probability function} \\ \sum_{x}^{\infty} e^{tx} f(x), & \text{for discrete probability function.} \end{cases}$$

Theorem: If $X_1, X_2, \dots X_n$ are independent random variables, then $M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \dots M_{X_n}(t)$.

Proof: By definition

$$M_{X_1+X_2+\dots+X_n}(t) = E\left[e^{t(X_1+X_2+\dots+X_n)}\right]$$

$$= E\left[e^{tX_1} \cdot e^{tX_2} \dots e^{tX_n}\right]$$

$$= E(e^{tX_1}) \cdot E(e^{tX_2}) \dots E(e^{tX_n})$$

$[\because X_1, X_2 \dots X_n \text{ are independent}]$

$$\boxed{M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \dots M_{X_n}(t)}$$

Theorem: If $U = \frac{X-a}{h}$, then $M_U(t) = e^{-\frac{at}{h}} \cdot M_X\left(\frac{t}{h}\right)$, a, h are constants.

Proof: By definition, $M_U(t) = E[e^{tU}]$

$$= E\left[e^{t\left(\frac{X-a}{h}\right)}\right]$$

$$= E\left[e^{\frac{tX}{h} - \frac{at}{h}}\right]$$

$$= e^{-\frac{at}{h}} \cdot E\left[e^{\left(\frac{t}{h}\right)X}\right]$$

$$\boxed{M_U(t) = e^{-\frac{at}{h}} M_X\left(\frac{t}{h}\right)} \quad \text{where } U = \frac{X-a}{h}$$

Find the m.g.f of the random Variable with the probability law $P(X=x) = q^{x-1}p$, $x = 1, 2, 3$

Solution: We know that

$$M_X(t) = E(e^{tx}) = \sum_{x=1}^{\infty} e^{tx} p(x)$$

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

$$= \sum_{x=1}^{\infty} e^{tx} q^{x} q^{-1} \cdot p = \sum_{x=1}^{\infty} (e^t q)^x \frac{p}{q}$$

$$= \frac{p}{q} \cdot qe^t \sum_{x=1}^{\infty} (qe^t)^{x-1}$$

$$= pe^t \left[1 + qe^t + (qe^t)^2 + \cdots \right]$$

$$= pe^t \left[1 - qe^t \right]^{-1} = \frac{pe^t}{1-qe^t}$$

$$\boxed{M_X(t) = \frac{pe^t}{1-qe^t}} \longrightarrow (1)$$

Differentiating (1) w.r.to 't' we get

$$\frac{d}{dt}\{M_X(t)\} = \frac{(1-qe^t)pe^t - pe^t(-qe^t)}{(1-qe^t)^2}$$

$$= \frac{pe^t - pqe^{2t} + pqe^{2t}}{(1-qe^t)^2}$$

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

$$\therefore \mu_1' \text{ (about origin)} = M_X'(0) = \frac{p}{(1-q)^2} \quad [\text{put } x=0 \text{ in } (2)]$$

$$= \frac{p}{p^2} = 1/p$$

Differentiating (2) w.r.to 't' we get

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

$$= \frac{(1-qe^t)\{(1-qe^t)pe^t + 2pqe^{2t}\}}{(1-qe^t)^4}$$

$$= pe^t + pqe^{2t} / (1-qe^{2t})^3$$

$$M_X''(t) = pe^t(1+qe^t) / (1-qe^{2t})^3 \longrightarrow (3)$$

$$\therefore \mu_2' \text{ (about origin)} = M_X''(0) = p(1+q) / (1-q)^3$$

$$\text{Mean} = \mu_1' = 1/p \text{ ; Variance} = \mu_2' - \mu_1'^2 = \frac{1+q}{p^2} - \frac{1}{p^2}$$

$$\boxed{\text{Variance} = q/p^2}$$

2. Find the m.g.f of the random variable whose moments are

$$\mu_r' = (r+1)!\,2^r$$

Solution: We know that the m.g.f in terms of moments is given by $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r'$

$$= \sum_{r=0}^{\infty} (r+1)!\,2^r \frac{t^r}{r!} \quad \left[\because \mu_r' = (r+1)!\,2^r\right]$$

$$= \sum_{r=0}^{\infty} \frac{(2t)^r (r+1)!}{r!}$$

(i.e) $M_X(t) = 1 + 2(2t) + 3(2t)^2 + \cdots$

$$\boxed{M_X(t) = (1-2t)^{-2}} \quad \left[\text{using } (1-x)^{-2} = 1+2x+3x^2+\cdots\right]$$

3. Find the m.g.f of the random variable 'X' having p.d.f

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

Solution: We know that

$$M_X(t) = \int_0^{\infty} e^{tx} f(x)\,dx \quad [\text{Here 'x' is a continuous variable}]$$

$$= \int_0^1 e^{tx} f(x)\,dx + \int_1^2 e^{tx} f(x)\,dx$$

$$= \int_0^1 e^{tx}\,x\,dx + \int_1^2 e^{tx}(2-x)\,dx$$

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

(using integration by parts)

$$= \frac{e^{2t}}{t^2} + \frac{1}{t^2} - \frac{2e^t}{t^2} = (e^t-1)^2/t^2$$

4. Find the m.g.f of a random variable 'x' having the p.d.f

$f(x) = 1/3, \quad -1 < x < 2$
$\quad\;\; = 0, \quad \text{otherwise}$

Solution: We know that the m.g.f for a continuous random variable 'X' is $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$

$$= \int_{-1}^{2} e^{tx} \cdot \tfrac{1}{3}\,dx$$

$$= \tfrac{1}{3}\left[\frac{e^{tx}}{t}\right]_{-1}^{2} = \tfrac{1}{3}\left[\frac{e^{2t} - e^{-t}}{t}\right]$$

Hence the problem.

habychov's inequality

If 'X' is a random variable with mean μ and variance σ^2 then for any positive number 'K' we have

$$P\{|X-\mu| \geq K\sigma\} \leq \frac{1}{K^2} \text{ (or) } P\{|X-\mu| < K\sigma\} \geq 1-\frac{1}{K^2}$$

Solution Let 'X' be a continuous random variable

By definition of variance, we have $\sigma^2 = \sigma_X^2 = E[X-E(X)]^2$

$$= E[X-\mu]^2 \qquad \mu = E(X)$$

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

$f(x)$ is p.d.f of X.

By defn. of expectation for continuous random variable

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

$$\sigma^2 \geq \int_{-\infty}^{\mu-K\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+K\sigma}^{\infty} (x-\mu)^2 f(x)\,dx \longrightarrow ①$$

In the first integral x takes values less than or equal to $\mu-K\sigma$. (ie) $x \leq \mu-K\sigma \Rightarrow \mu-x \leq K\sigma \longrightarrow ②$

In the second integral x takes values greater than or equal to $\mu+K\sigma$ (ie) $x \geq \mu+K\sigma \Rightarrow x-\mu \geq K\sigma$

(ie) $\mu-x \leq -K\sigma \longrightarrow ③$

Substituting ② & ③ in ① we get

$$\sigma^2 \geq \int_{-\infty}^{\mu-K\sigma} K^2\sigma^2 f(x)\,dx + \int_{\mu+K\sigma}^{\infty} K^2\sigma^2 f(x)\,dx$$

$$= K^2\sigma^2 \left[\int_{-\infty}^{\mu-K\sigma} f(x)\,dx + \int_{\mu+K\sigma}^{\infty} f(x)\,dx\right]$$

$$= K^2\sigma^2 \{P(X \leq \mu-K\sigma) + P(X \geq \mu+K\sigma)\}$$

$$= K^2\sigma^2 \{P(X-\mu \leq -K\sigma) + P(X-\mu \geq K\sigma)\}$$

$$= K^2\sigma^2 P\{|X-\mu| \geq K\sigma\}$$

(ie) $\sigma^2 \geq K^2\sigma^2 P\{|X-\mu| \geq K\sigma\}$

(ie) $P\{|X-\mu| \geq K\sigma\} \leq \frac{1}{K^2}$

Since total probability is 1, we have

$P\{|X-\mu| \geq K\sigma\} + P\{|X-\mu| < K\sigma\} = 1$

(ie) $P\{|X-\mu| < K\sigma\} = 1 - P\{|X-\mu| \geq K\sigma\}$

$$\geq 1 - \frac{1}{K^2}$$

Hence the proof.

Example 1 A random variable X has a mean $\mu = 12$ and a variance $\sigma^2 = 9$ and an unknown probability distribution. Find $P(6 < x < 18)$.

Solution: Given $\mu = 12$, $\sigma^2 = 9 \Rightarrow \sigma = 3$.

By chebychev's inequality, we have $P\{|X-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}$

$$P\{-k\sigma < X-\mu < k\sigma\} \geq 1 - \frac{1}{K^2}$$

(ie) $P\{\mu - k\sigma < X < \mu + k\sigma\} \geq 1 - \frac{1}{K^2} \longrightarrow (1)$

Now $P\{6 < X < 18\} = 1 - 1/K^2 \longrightarrow (2)$

(ie) $\left.\begin{array}{l}\mu - k\sigma = 6 \\ \mu + k\sigma = 18\end{array}\right\} \Rightarrow \left.\begin{array}{l}12 - 3K = 6 \\ 12 + 3K = 18\end{array}\right\} \longrightarrow (3)$

$2\mu = 24 \Rightarrow \mu = 12$ and then $\sigma = 3$

Solving (3), we get, $K = 2$. Substituting $K = 2$ in (2) we get

$$\therefore P\{6 < X < 18\} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$$

Hence the problem.

(2) A random variable 'x' has a mean 10 and a variance 4 and an unknown probability distribution. Find the value of 'c' such that $P\{|X-10| \geq c\} \leq 0.04$.

Solution: Given $\mu = 10$, $\sigma^2 = 4 \Rightarrow \sigma = 2$.

By chebychev's inequality, we have

$$P\{|X-\mu| \geq k\sigma\} \leq 1/k^2 \longrightarrow (1)$$

Given $P\{|X-10| \geq c\} \leq 0.04 \longrightarrow (2)$

Comparing (1) & (2) we get

$$\frac{1}{k^2} = 0.04 \text{ and } k\sigma = c.$$

$$\Rightarrow \frac{1}{k^2} = \frac{4}{100} \Rightarrow k^2 = \frac{100}{4} \Rightarrow K = \frac{10}{2} = 5$$

$$\Rightarrow k\sigma = c \Rightarrow c = 2k = 2(5) = 10 \Rightarrow \boxed{c = 10}$$

Hence the problem.

Unit - IV

Correlation Analysis

Correlation: If the change in one variable affects a change in the other variable, the variables are said to be correlated.

If the two variables deviate in the same direction (ie) if the increase (or decrease) in one results in a corresponding increasing (or decrease) in the other, correlation is said to be direct (or) positive correlation.

But if they constantly deviate in opposite direction (ie) if increase (or decrease) in one results in corresponding decrease (or increase) in the other, correlation is said to be negative.

Example: The correlation between (a) the heights and weights of a group of persons (b) income and expenditure is positive. The correlation between (a) price and demand of a commodity and (b) the correlation between volume and pressure of a perfect gas is negative.

Karl - pearson's coefficient of correlation:

Correlation coefficient between two random variables X and Y, usually denoted by $r(X,Y)$ is a numerical measure of linear relationship between them and is defined as
$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

where
$$\operatorname{cov}(X,Y) = \frac{1}{n}\Sigma XY - \bar{X}\bar{Y}$$
$$\sigma_X = \sqrt{\frac{1}{n}\Sigma X^2 - \bar{X}^2}$$
and
$$\sigma_Y = \sqrt{\frac{1}{n}\Sigma Y^2 - \bar{Y}^2}$$

n is the number of items in the given data.

Note: Correlation co-efficient cannot exceed unity numerically (ie) $-1 \le r_{xy} \le 1$.

Example : 1

Calculate the correlation coefficient for the following heights (in inches) of fathers X and their Sons Y.

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Solution: Method : I

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
544	552	37560	37028	38132

Now $\bar{X} = \frac{544}{8} = 68$

$\bar{Y} = \frac{552}{8} = 69$

$\bar{X}\bar{Y} = 68 \times 69$
$= 4692$

$\sigma_X = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}$
$= \sqrt{\frac{37028}{8} - 4624}$
$= 2.121$

$\sigma_Y = \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}$

$= \sqrt{\frac{38132}{8} - 4761} = 2.345$

$\therefore r(X,Y) = \frac{\frac{1}{n}\sum XY - \bar{X}\bar{Y}}{\sigma_X \cdot \sigma_Y} = \frac{\frac{1}{8} \times 37560 - 4692}{2.121 \times 2.345}$

$= 0.6030$

(ii) $r(X,Y) = r(u,v)$

where $u = \frac{X-a}{h}$; $v = \frac{Y-b}{K}$

where a and b are some arbitrary constants usually the mid values of the given data X and Y respectively

Method : 2

Now $\bar{u} = \frac{\sum u}{n} = \frac{0}{8} = 0$

$\bar{v} = \frac{\sum v}{n} = \frac{8}{8} = 1$

$\mathrm{cov}(X,Y) = \mathrm{cov}(u,v)$
$= \frac{\sum uv}{n} - \bar{u}\bar{v} = \frac{24}{8} - 0 = 3$

$\sigma_u = \sqrt{\frac{\sum u^2}{n} - \bar{u}^2} = 2.121$ & $\sigma_v = \sqrt{\frac{\sum v^2}{n} - \bar{v}^2} = 2.345$

X	Y	$u = X - 68$	$v = Y - 68$	uv	u^2	v^2
65	67	-3	-1	3	9	1
66	68	-2	0	0	4	0
67	65	-1	-3	3	1	9
67	68	-1	0	0	1	0
68	72	0	4	0	0	16
69	72	1	4	4	1	16
70	69	2	1	2	4	1
72	71	4	3	12	16	9
		0	8	24	36	52

$$\therefore r(X,Y) = r(u,v) = \frac{\operatorname{cov}(u,v)}{\sigma_u \cdot \sigma_v} = \frac{3}{2.121 \times 2.345} = 0.6031$$

2. Find the correlation coefficient for the following data

X	10	14	18	22	26	30
Y	18	12	24	6	30	36

Solution:

x	y	$u = \frac{X-22}{4}$	$v = \frac{Y-24}{6}$	uv	u^2	v^2
10	18	-3	-1	3	9	1
14	12	-2	-2	4	4	4
18	24	-1	0	0	1	0
22	6	0	-3	0	0	9
26	30	1	1	1	1	1
30	36	2	2	4	4	4
		-3	-3	12	19	19

Now $\bar{u} = \frac{\Sigma u}{n} = \frac{-3}{6} = -0.5$

$\bar{v} = \frac{\Sigma v}{n} = -3/6 = -0.5$

$\therefore \operatorname{cov}(u,v) = \frac{\Sigma uv}{n} - \bar{u}\,\bar{v} = 7/4$

$\sigma_u = \sqrt{\frac{\Sigma u^2}{n} - \bar{u}^2} = 1.708$

$\sigma_v = \sqrt{\frac{\Sigma v^2}{n} - \bar{v}^2} = 1.708$

$\therefore r(X,Y) = r(u,v) = \frac{\operatorname{cov}(u,v)}{\sigma_u \cdot \sigma_v}$

$= 0.6$

$$\boxed{\therefore r(X,Y) = 0.6}$$

Hence the problem.

Rank Correlation

Let (x_i, y_i), $i = 1, 2, \ldots n$ be the ranks of 'i' individuals in two characteristics A and B respectively. Pearsonian co-efficient of correlation between the ranks x_i's and y_i's is called the rank correlation co-efficient between the characteristics A and B for that group of individuals and is given by

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}, \text{ where } d_i = x_i - y_i$$

Example: 1 Find the rank correlation coefficient from the following data

Rank in X	1	2	3	4	5	6	7
Rank in Y	4	3	1	2	6	5	7

Solution:

X	Y	$d_i = x_i - y_i$	d_i^2
1	4	-3	9
2	3	-1	1
3	1	2	4
4	2	2	4
5	6	-1	1
6	5	1	1
7	7	0	0
		0	20

Rank correlation coefficient

$$r(X,Y) = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 0.6429.$$

Hence the problem.

Example: 2 10 competitors in a musical test were ranked by the 3 judges x, y, z in the following order.

	A	B	C	D	E	F	G	H	I	J
Rank by x	1	6	5	10	3	2	4	9	7	8
Rank by y	3	5	8	4	7	10	2	1	6	9
Rank by z	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common likings of music

Solution

x_i	y_i	z_i	$d_1 = x_i - y_i$	$d_2 = y_i - z_i$	$d_3 = x_i - z_i$	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
						200	214	60

The rank correlation between x & y is

$$r_1(x,y) = 1 - \frac{6\sum d_1^2}{n(n^2-1)} = -0.212$$

The rank correlation between y and z is

$$r_2(x,y) = 1 - \frac{6\sum d_2^2}{n(n^2-1)} = -0.296$$

The rank correlation between x and z is

$$r_3(x,z) = 1 - \frac{6\sum d_3^2}{n(n^2-1)} = 0.636$$

Since the rank correlation between x and z is maximum and also positive, we conclude that the pair of judges x and z has the nearest approach to common likings in music.

Repeated Ranks:

In the correlation formula, we add the factor $\frac{m(m^2-1)}{12}$ to $\sum d^2$ where m is the number of items an item is repeated. This correction factor is to be added for each repeated value.

Example: 1 A sample of 12 fathers and their eldest sons have the following data about their heights in inches

Fathers	65	63	67	64	68	62	70	66	68	67	69	71
Sons	68	66	68	65	69	66	68	65	71	67	68	70

Calculate the rank correlation coefficient.

Solution:

Fathers	Sons	Rank of x_i	Rank of y_i	$d_i = x_i - y_i$	d_i^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	-1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	-3.5	12.25
66	65	8	11.5	-3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	-1.5	2.25
69	68	3	5.5	-2.5	6.25
71	70	1	2	-1	1
					72.5

Correction Factors:

In X Series 68 is repeated twice

$\therefore C.F = 2(2^2-1)/12 = \frac{1}{2}$

In X Series 67 is repeated twice

$\therefore C.F = 2(2^2-1)/12 = \frac{1}{2}$

In Y Series 68 is repeated four times

$\therefore C.F = 4(4^2-1)/12 = 5$

In Y Series 66 is repeated twice

$\therefore C.F = \frac{2(2^2-1)}{12} = \frac{1}{2}$

In Y Series 65 is repeated twice $\therefore C.F = \frac{2(2^2-1)}{12} = \frac{1}{2}$

$\therefore$ Rank Correlation coefficient

$$r(x,y) = 1 - \frac{6[72.5+0.5+0.5+5+0.5+0.5]}{12(144-1)}$$

$$= 0.722$$

Regression:

Regression is a mathematical measure of the average relationship between two or more variables interms of the original limits of the data.

Lines of Regression:

If the variables in a bivariate distribution are related we will find that the points in the scattered diagram will cluster around some curve called the curve of regression.

If the curve is a straight line, it is called the line of regression and there is said to be linear regression between the variables, otherwise regression is said to be curvilinear.

The line of regression of y on x is given by

$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

where r is the correlation coefficient σ_y and σ_x are standard deviations.

The line of regression of x on y is given by

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

Note: Both the lines of regression passes through $(\bar{x}, \bar{y})$

Angle between two lines of Regression

If the equations of lines of regression of y on x and x on y are $y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x})$ and

$$x - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

The angle 'θ' between the two lines of regression is given by $\tan\theta = \frac{1 - r^2}{r}\left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}\right)$

Note: 1 If $r = 0$, we get $\tan\theta = \infty \Rightarrow \theta = \pi/2$

$\therefore$ when $r = 0$ the lines of regression are perpendicular to each other.

Note: 2 If $r = \pm 1$, then $\tan\theta = 0 \Rightarrow \theta = 0$ or π

$\therefore$ when $r = \pm 1$, the two regression lines are parallel to each other (or) coincide.

Note: 3 when $r = 0$, the two variables X and Y are uncorrelated.

Note: 4 when $r = \pm 1$, the correlation between X and Y is said to be perfect.

Regression Coefficients

Regression coefficient of y on x $= r\frac{\sigma_y}{\sigma_x} = b_{yx}$ — (1)

Regression coefficient of x on y $= r\frac{\sigma_x}{\sigma_y} = b_{xy}$ — (2)

From (1) & (2) we get

$$r\frac{\sigma_y}{\sigma_x} \cdot r\frac{\sigma_x}{\sigma_y} = b_{yx} \times b_{xy}$$

$$r^2 = b_{xy} \times b_{yx} \Rightarrow r = \pm\sqrt{b_{xy} \times b_{yx}}$$

(ie) correlation coefficient $r = \pm\sqrt{b_{xy} \times b_{yx}}$

The regression coefficients b_{yx} and b_{xy} can be easily obtained by using the following formula.

Note :1 $b_{yx} = \Sigma(x-\bar{x})(y-\bar{y}) / \Sigma(x-\bar{x})^2$

$b_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) / \Sigma(y-\bar{y})^2$

Note : 2 Regression coefficients are independent of change of origin but not of scale.

(ie) $$b_{yx} = b_{vu} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma u^2 - (\Sigma u)^2}$$

and $$b_{xy} = b_{uv} = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{n\Sigma v^2 - (\Sigma v)^2}$$

where $u = x-a$, $v = y-b$.

Example :1 From the following data, find

(i) The two regression equations

(ii) The coefficient of correlation between 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 : Here $\bar{x} = \frac{\Sigma x}{n} = \frac{320}{10} = 32$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{380}{10} = 38$$

coefficient of regression of y on x is

$$b_{yx} = \Sigma(x-\bar{x})(y-\bar{y}) / \Sigma(x-\bar{x})^2 = -93/140 = -0.6643$$

x	y	$x-\bar{x}$ $x-32$	$y-\bar{y}$ $y-38$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
31	36	-1	-2	1	4	2
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
320	380	0	0	140	398	-93

coefficient of regression of x on y is

$$b_{xy} = \Sigma(x-\bar{x})(y-\bar{y}) / \Sigma(y-\bar{y})^2 = -93/398 = -0.2337$$

Equation of the line of regression of x on y is

$$(x-\bar{x}) = b_{xy}(y-\bar{y})$$

(i) $x - 32 = -0.2337(y-38)$

$$= -0.2337y + 0.2337 \times 38$$

$$\boxed{x = -0.2337y + 40.8806}$$

Equation of the line of regression of y on x is

$$y - \bar{y} = b_{yx}(x-\bar{x})$$

$$y - 38 = -0.6643(x-32)$$

$$y = -0.6643x + 38 + 0.6643 \times 32$$

$$\boxed{y = -0.6643x + 59.2576}$$

Coefficient of correlation $r^2 = b_{yx} \times b_{xy}$

$$= -0.6643 \times (-0.2337)$$

$$r^2 = 0.1552$$

$$r = \pm\sqrt{0.1552}$$

$$= \pm 0.394$$

Now we have to find the most likely marks in Statistics (y) when marks in Economics (x) are 30. We use the line of regression of y on x.

(ie) $y = -0.6643x + 59.2575$

Put $x = 30$, we get $y = -0.6643 \times 30 + 59.2536$

$$y \simeq 39$$

Example: The two lines of regression are

$8x - 10y + 66 = 0$ — ① $40x - 18y - 214 = 0$ → ②

The variance of x is 9. Find (i) the mean values of x & y

(ii) correlation coefficient between x and y.

Solution (i) Since both the lines of regression passes through the mean values $\bar{x}$ and $\bar{y}$, the point $(\bar{x}, \bar{y})$ must satisfy the two given regression lines

$$8\bar{x} - 10\bar{y} = -66 \longrightarrow ③$$
$$40\bar{x} - 18\bar{y} = 214 \longrightarrow ④$$

③ × 5 ⇒ $40\bar{x} - 50\bar{y} = -330 \longrightarrow ⑤$

$$40\bar{x} - 18\bar{y} = 214$$

④ − ⑤ ⇒ $32\bar{y} = 544 \Rightarrow \boxed{\bar{y} = 17}$

Substituting $\bar{y} = 17$ in ③ we get,

$$8\bar{x} - 10\times17 = -66 \Rightarrow \boxed{\bar{x} = 13}$$

Hence the mean values are given by $\bar{x} = 13$, $\bar{y} = 17$.

(ii) Let us suppose that equation ① is the equation of line of regression of y on x and ② is the equation of the line regression of x on y. we get, after rewriting ① & ②

$$10y = 8x + 66 \Rightarrow y = \frac{8}{10}x + \frac{66}{10} \qquad \therefore \boxed{b_{yx} = \frac{8}{10}}$$

$$40x = 18y + 214 \Rightarrow x = \frac{18}{40}y + \frac{214}{40} \qquad \therefore \boxed{b_{xy} = \frac{18}{40}}$$

$$\therefore r^2 = b_{xy} \times b_{yx}$$
$$= \frac{18}{40} \times \frac{8}{10} = \frac{9}{25}$$

$$\therefore r = \pm\frac{3}{5} = \pm 0.6$$

Since both the regression coefficients are positive, r must be positive.

$r = 0.6$

Hence the problem.

DISCRETE DISTRIBUTIONS

Binomial Distribution

The probability of x successes in n trials is given by $nC_x p^x q^{n-x}$ (i.e) $P(x \text{ successes}) = nC_x p^x q^{n-x}$

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

Here $nC_x p^x q^{n-x}$ is the $(x+1)^{th}$ term in the expansion of $(q+p)^n$

$$\left[\because (q+p)^n = q^n + nC_1 q^{n-1} p + nC_2 q^{n-2} p^2 + \dots + nC_x q^{n-x} p^x + \dots\right]$$

which is a Binomial series and hence the distribution is called Binomial distribution.

Mean and Variance of the Binomial Distribution

We know that, for discrete probability distribution mean is given by $\mu_1' = E(X)$ $\quad [\because E(X) = \bar{x}]$

$$= \sum_{x=0}^{n} x P(x) \quad [P(x) \text{ is p.d.f}]$$

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

$$= 0 \cdot q^n + 1 \cdot nC_1 q^{n-1} p + 2 \cdot nC_2 q^{n-2} p^2 + \dots + np^n$$

$$= np\left[q^{n-1} + n-1C_1 q^{n-2} p^1 + \dots + p^{n-1}\right]$$

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

Hence mean of the binomial distribution is $\bar{x} = np$

Now, $\mu_2' = \sum_{x=0}^{n} x^2 p(x) = \sum_{x=0}^{n} x^2 \cdot nC_x p^x q^{n-x}$

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

$$= \sum_{x=0}^{n} x \, nC_x p^x q^{n-x} + \sum_{x=0}^{n} x(x-1) nC_x p^x q^{n-x}$$

$$= np + \sum_{x=0}^{n} n(n-1)(n-2)C_{(x-2)} p^2 \cdot p^{x-2} q^{n-x}$$

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

$$\boxed{\mu_2' = np[q + np]}$$

$\therefore$ Variance $(\mu_2) = \mu_2' - \mu_1'^2 = np(q+np) - n^2p^2 = npq$

Standard deviation $= \sqrt{npq}$

Moment generating function (m.g.f) of binomial distribution about origin

We know that the moment generating function of random variable X about origin whose probability function $f(x)$ is given by $M_x(t) = \sum_{x=0}^{n} e^{tx} f(x)$

Here, let 'x' be a random variable which follows binomial distribution

Then its m.g.f about origin is given by

$$M_x(t) = E[e^{tx}] = \sum_{x=0}^{n} e^{tx} f(x)$$

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

$$= \sum_{x=0}^{n} (e^t)^x p^x nC_x q^{n-x}$$

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

$$= q^n + nC_1 q^{n-1}(pe^t)^1 + nC_2 q^{n-2}(pe^t)^2 + \cdots$$

$$\boxed{M_x(t) = (q+pe^t)^n}$$

Moment generating function (m.g.f) of a binomial distribution about mean (np)

We know that the m.g.f of a random variable X about any point 'a' is $M_X(t)$ (about $X=a$) $= E\left[e^{t(X-a)}\right]$

Here 'a' is mean of the binomial distribution

(ie) $M_X(t)$ (about $X = np$) $= E\left[e^{t(X-np)}\right]$

$$= e^{-tnp} E[e^{tX}]$$

$$= e^{-tnp} \text{ m.g.f of X about origin}$$

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

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

$$= \{qe^{-tp} + pe^{tq}\}^n$$

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

$$\left[\because e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots\right]$$

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

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

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

Now μ_2 = coefficient of $\frac{t^2}{2!} = npq$

(i.e) Variance $= npq$

μ_3 = coefficient of $\frac{t^3}{3!} = npq(q-p)$ and so on.

Recurrence Relation for the moments of Binomial Distribution

We know that $\mu_r = E[X - E(X)]^r$

$$= \sum_{x=0}^{n} (x-np)^r \binom{n}{x} p^x q^{n-x}$$

Differentiating w.r.to p,

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} \binom{n}{x} \left[p^x q^{n-x} r(x-np)^{r-1}(-n) + (x-np)^r q^{n-x} x p^{x-1} + (x-p)^r p^x (n-x) q^{n-x-1}(-1)\right] \quad [\because q = 1-p]$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left[-nr(x-np)^{r-1} p^x q^{n-x} + (x-np)^r \left\{x p^{x-1} q^{n-x} - (n-x) p^x q^{n-x-1}\right\}\right]$$

$$= -nr \sum_{x=0}^{n} \binom{n}{x} (x-np)^{r-1} p^x q^{n-x} + \sum_{x=0}^{n} \binom{n}{x} (x-np)^r p^x q^{n-x} \left\{\frac{x}{p} - \frac{n-x}{q}\right\}$$

$$= -nr \sum_{x=0}^{n} (x-np)^{r-1} p(x) + \sum_{x=0}^{n} (x-np)^r p(x) \left(\frac{x-np}{pq}\right)$$

$$\left[\because p(x) = \binom{n}{x} p^x q^{n-x}\right]$$

$$= -nr \sum_{x=0}^{n} (x-np)^{r-1} p(x) + \frac{1}{pq} \sum_{x=0}^{n} (x-np)^{r+1} p(x)$$

$$\therefore \frac{d\mu_r}{dp} = -nr\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$$

(i.e) $$\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right] \longrightarrow (1)$$

This gives the recurrence relation for the moments of Binomial distribution.

Put $n=1$, in ①, we get $\mu_2 = pq\left[n\mu_0 + \frac{d\mu_1}{dp}\right] = npq$ $[\because \mu_1 = 0, \mu_0 = 1]$

Put $n=2$ in ① we get $\mu_3 = pq\left[2n\mu_1 + \frac{d\mu_2}{dp}\right] = pq\frac{d\mu_2}{dp}$ $[\because \mu_1 = 0]$

$= pq\frac{d(npq)}{dp}$

$$\boxed{\mu_3 = npq(q-p)}$$

Additive property of Binomial Distribution

Let X and Y be two independent binomial random variables

Then $M_X(t) = (q_1 + p_1e^t)^{n_1}$, $M_Y(t) = (q_2 + p_2e^t)^{n_2}$

Now $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ $[X, Y - \text{independent}]$

$= (q_1 + p_1e^t)^{n_1} \cdot (q_2 + p_2e^t)^{n_2} \longrightarrow ①$

$\neq (q + pe^t)^n$ $[\because$ m.g.f is unique$]$

Equation ① cannot be expressed in the form $(q+pe^t)^n$

Hence $X+Y$ is not a binomial variate. Hence the sum of two independent binomial variates is not a binomial variate.

1. The mean and variance of a binomial distribution are 4 and $\frac{4}{3}$. Find $P(X \geq 1)$.

Solution: we know that mean of a binomial distribution is np and variance is npq. Given $np = 4 \longrightarrow ①$, $npq = \frac{4}{3} \longrightarrow ②$

$①/② \Rightarrow \frac{\cancel{np}}{\cancel{np}q} = \frac{4}{4/3} \Rightarrow \frac{1}{q} = \frac{3}{1} \Rightarrow \boxed{q = 1/3}$ $\boxed{P(x) = nC_x p^x q^{n-x}}$

But $p+q=1 \Rightarrow p = 1 - 1/3 = 2/3$

Substituting $p = 2/3$ in ① we get $n = \frac{4}{p} = \frac{4}{2/3} = 6$

Now $P(X \geq 1) = 1 - P(0) = 1 - nC_0 p^0 q^{6-0} = 1 - q^6 = 1 - (1/3)^6 = 0.998$

2. In a large consignment of electric bulbs 10% are defective. A random sample of 20 is taken for inspection. Find the probability that (i) All are good bulbs (ii) At most there are 3 defective bulbs (iii) Exactly there are three defective bulbs

Solution: Here $p = 10/100 = 0.1$, $q = 1-p = 0.9$, $n = 20$

(i) $P(\text{all are good bulbs}) = P(\text{none are defective}) = P(0)$

$= nC_0 p^0 q^{n-0}$

$= 20C_0 (0.1)^0 (0.9)^{20}$

$= 0.1216$

(ii) $P(\text{at most there are 3 defective bulbs})$

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

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

$= 0.1215 + 0.27 + 0.285 + 0.19$

$= 0.8666$

(iii) $P(\text{exactly 3 defective bulbs}) = P(3) = nC_3 p^3 q^{n-3}$

$= 20C_3 (0.1)^3 (0.9)^{17} = 0.19$

With the usual notation find 'p' for a binomial random variable 'X' if $n=6$ and if $9P(X=4)=P(X=2)$.

Solution: We know that for binomial random variable 'X':

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

Given $9P(X=4)=P(X=2)$

$$9\times 6C_4\, p^4q^2 = 6C_2\, p^2q^4 \Rightarrow 9p^2=q^2=(1-p)^2$$

$$\Rightarrow 9p^2 = 1+p^2-2p \Rightarrow 8p^2+2p-1=0$$

$$p=\frac{-2\pm\sqrt{4+32}}{16}$$

$$=\tfrac{1}{4}=0.25$$

$$\boxed{p=0.25,\ \therefore q=0.75}$$

Hence the proof

POISSON DISTRIBUTION

The Probability function of a random variable "x" which follows Poisson distribution is given by.

$$P[x=x]=\frac{e^{-\lambda}\lambda^x}{x!},\ x=0,1,2,\ldots\infty$$

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

Moment Generating Function of the Poisson Distribution

We Know that the m.g.f of random Variable 'X' is given by

$$M_X(t)=\sum_{x=0}^{n} e^{tx} f(x) \quad [f(x) \text{ is p.d.f of Poisson Distribution}]$$

$$=\sum_{x=0}^{n} e^{t}\left(\frac{e^{-\lambda}\lambda^x}{x!}\right) \quad \left[f(x)=\frac{e^{-\lambda}\lambda^x}{x!}\right].$$

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

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

$$=e^{-\lambda}\cdot e^{\lambda e^t}=e^{\lambda(e^t-1)}$$ Hence $M_X(t)=e^{\lambda(e^t-1)}$

$\therefore$ Moment generating function of the random Variable 'X' is

$$M_X(t)=e^{\lambda(e^t-1)}$$

To find mean and variance of the poisson Distribution

We Know that, for discrete probability distribution, mean is given by

$$\mu_1' = E(X) = \sum_{x=0} x \cdot p(x)$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{n} \frac{x e^{-\lambda} \lambda \cdot \lambda^{x-1}}{x!}$$

$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

Hence the mean of the poisson distribution is λ.

Now $\mu_2' = E(x^2) = \sum_{x=0}^{\infty} x^2 p(x) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}$

$$= \sum_{x=0}^{\infty} (x(x-1)+x) \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{x e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda \quad \left[\text{Using } \frac{1}{n!} = 0 \text{ when 'n' is negative}\right]$$

$$= e^{-\lambda}\lambda^2 \left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right] + \lambda = e^{-\lambda} e^{\lambda} \cdot \lambda^2 + \lambda = \lambda^2 + \lambda$$

Variance $\mu_2 = \mu_2' - \mu_1'^2 = (\lambda^2 + \lambda) - \lambda^2$

$\boxed{\text{Variance} = \lambda}$. Hence $\boxed{\text{mean} = \text{Variance} = \lambda}$

Recurrence Relation for the moments of Poisson Distribution.

We know that $\mu_r = E[X - E(X)]^r$

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

Differentiating w.r.to 'λ' we get

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} \frac{1}{x!}\left[r(x-\lambda)^{r-1}(-1) \cdot e^{-\lambda}\lambda^x + (x-\lambda)^r e^{-\lambda}(-1)\lambda^x + (x-\lambda)^r e^{-\lambda} \cdot x\lambda^{x-1}\right]$$

$$= -r\sum_{x=0}^{\infty} (x-\lambda)^{r-1} \frac{e^{-\lambda}\lambda^x}{x!} + \frac{1}{\lambda}\sum_{x=0}^{\infty} (x-\lambda)^{r+1} \cdot \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\frac{d\mu_r}{d\lambda} = -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1} \quad \text{(ie)} \quad \mu_{r+1} = r\lambda\mu_{r-1} + \lambda \frac{d\mu_r}{d\lambda} \rightarrow (1)$$

Putting $r = 1, 2, 3,$ we get $\mu_2 = r\mu_0 + \lambda \frac{d\mu_1}{d\lambda} = \lambda$

$$\mu_3 = 2\lambda\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda \qquad [\because \mu_1 = 0, \mu_0 = 1]$$

$$\mu_4 = 3\lambda\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda \qquad [\because \mu_1 = 0, \mu_2 = \lambda]$$

Additive property of Independent poisson Variates

Sum of independent poisson Variates is also a poisson Variate.

(ii) if X_i $(i=1,2,\dots n)$ are independent poisson variates with parameters λ_i $(i=1,2\dots n)$ then $\sum_{i=1}^{n} X_i$ is also a poisson variate with parameter $\sum_{i=1}^{n} \lambda_i$.

Proof :- We know that the m.g.f of the poisson variate X_i is given by $M_{X_i}(t) = e^{\lambda_i(e^t-1)}$, $i=1,2\dots n$

(ie) $M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t)\cdot M_{X_2}(t)\dots M_{X_n}(t)$

$$= e^{\lambda_1(e^t-1)}\cdot e^{\lambda_2(e^t-1)}\dots e^{\lambda_n(e^t-1)}$$

$$= e^{(\lambda_1+\lambda_2+\dots+\lambda_n)(e^t-1)}$$

which gives the m.g.f of a poisson variate with parameter $\lambda_1+\lambda_2+\dots+\lambda_n$.

Examples :-

1. If X is a poisson variate $P(X=2) = 9P(X=4) + 90P(X=6)$ find (i) mean of X (ii) variance of X.

Soln :- $P(X=x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $x=0,1,2\dots$

Given $P(X=2) = 9P(X=4) + 90P(X=6)$

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

$$= e^{-\lambda}\lambda^2\left(\frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!}\right)$$

$$\frac{1}{2} = \frac{9\lambda^2}{4!} + \frac{90\lambda^4}{6!} \implies \lambda^4+3\lambda^2-4=0.$$

$$\lambda^2 = \frac{-3\pm\sqrt{9+16}}{2} = 1 \text{ (or) } -1$$

$$\lambda = \pm 1 \text{ (or) } \lambda = \pm i$$

Mean $=\lambda = 1$, Variance $=\lambda = 1 \implies \therefore$ standard deviation $=1$.

---

2. A manufacturer of cotterpins knows that 5% of his product is defective. If he sells cotterpins in boxes of 100 and guarantees that not more than 10 pins will be defective. What is the approximate probability that a box will fail to meet the guaranteed quality?

Soln :- Given $n=100$, $p = 5\% = \frac{5}{100} = 0.05$

$\therefore$ Mean $\lambda = n\times p = 100\times 0.05 = 5$

The poisson distribution is $p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-5}5^x}{x!}$

Now p (a box will fail to meet the guaranteed quality)

$= p(x>10) = 1-p(x\le 10) = 1-[p(0)+p(1)+ \cdots + p(10)]$

$= 1-e^{-5}\left[1+\frac{5}{1!}+\frac{5^2}{2!}+\frac{5^3}{3!}+ \cdots +\frac{5^{10}}{10!}\right]$

$= 1-e^{-5}[146.36] = 1-0.9863 = 0.014$

3. Find the probability that atmost 5 defective fuses will be found in a box of 200 fuses if experiences shows that 2% of such fuses are defective

Soln: Given $n=200$, $p=2\% = \frac{2}{100} = 0.02$

$\therefore$ Mean $\lambda = n \times p = 200 \times 0.02 \Rightarrow \lambda = 4$

The poisson distribution is $p(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-4}4^x}{x!}$

(i.e) p(of x defective bulbs) $= \frac{e^{-4}4^x}{x!}$

Now p(atmost 5 defective fuses) $= P(x\le 5)$

$= p(0)+p(1)+p(2)+p(3)+p(4)+p(5)$

$= e^{-4}[42.866]$

$= 0.785$

4. Six coins are tossed 6400 times. Using the poisson distribution, what is the approximate probability of getting six heads 10 times.

Soln: Given $n=6400$

Probability of getting one head with one coin $= 1/2$

$\therefore$ The probability of getting six heads with six coins $= (1/2)^6 = 1/64$

$\therefore$ Mean $\lambda = n \times p = 6400 \times \frac{1}{64} = 100$

The poisson distribution is $p(x) = \frac{e^{-100}(100)^x}{x!}$

(i.e) P(of getting x heads) $= \frac{e^{-100}(100)^x}{x!}$

$\therefore$ probability of getting six heads in 10 times

$= \frac{e^{-\lambda}\lambda^{10}}{\lambda!} = \frac{e^{-100}(100)^{10}}{(100)!}$

5. A manufacturer knows that the condensers he makes contain on the average 1% of defectives. He packs them in boxes of 100. What is the probability that a box picked at random will contain 3 or more faulty condensers.

Solution: Given $p = 1\% = \frac{1}{100} = 0.01$

$\lambda = n \times p = 100 \times 0.01 = 1$

The poisson distribution is $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

(ii) p(of x faulty condensers) $= \frac{e^{-1}(1)^x}{x!}$

P(3 or more faulty condensers) $= p(3)+p(4)+ \quad +p(100)$

$= 1 - \{p(0)+p(1)+p(2)\}$

$= 1 - e^{-1}\left[1+1+\frac{1}{2}\right] = 1 - e^{-1}\times 2.5$

$= 1 - 0.3679\times 2.5$

$= 0.08025$

CONTINUOUS DISTRIBUTIONS

Normal Distribution

A random variable X is said to follow normal distribution with mean μ and variance σ^2 if its density function is given by the probability law

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

* The total area bounded by the above curve is 1.

$$\text{Area} = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{(x-\mu)^2/2\sigma^2} dx$$

Let $z = \frac{x-\mu}{\sigma}$ when $x = x_1$, $z_1 = \frac{x_1-\mu}{\sigma}$

$\sigma dz = dx$ when $x = x_2$, $z_2 = \frac{x_2-\mu}{\sigma}$

$$P(z_1 < z < z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz = \phi(z)$$

The integral $\frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz$ is called the probability integral.

$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-z^2/2} dz$ is called the standard normal curve and it is bell shaped and symmetrical about the line $z=0$.

Moment Generating function of Normal Distribution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

Put $z = \frac{x-\mu}{\sigma}$ when $x = -\infty$, $z = -\infty$

$x = \infty$, $z = \infty$

$\therefore \sigma dz = dx$

$$M_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma z+\mu)} \cdot e^{-z^2/2} \sigma dz$$

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

$$e^{\mu t + \frac{t^2\sigma^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2} dz$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du, \quad u = z - \sigma t$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \times 2 \int_{0}^{\infty} e^{-u^2/2} du$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot 2 \cdot \frac{\sqrt{\pi}}{2} \quad \left[\because \int_0^{\infty} e^{-u^2/2} du = \frac{\sqrt{\pi}}{2} \right]$$

$$M_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$$

Moments ~~Abo~~ of Normal Distribution

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

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

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

Put $z = \frac{x-\mu}{\sigma} \Rightarrow \sigma z = x - \mu \Rightarrow \sigma dz = dx$

$$\therefore \mu_{2n+1} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot e^{-\frac{(\sigma z)^2}{2\sigma^2}} \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot e^{-z^2/2} dz = \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \cdot e^{-z^2/2} dz$$

$$= 0$$

[$\because$ The integrand is an odd function of z]

Even order moments about mean are given by

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x) dx = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \Gamma(n + 1/2)$$

$$\left[\because \Gamma n = \int_0^{\infty} e^{-x} x^{n-1} dx \right]$$

Changing n to n-1, we get

$$\mu_{2n-2} = \frac{2^{n-1}}{\sqrt{\pi}} \sigma^{2n-2} \Gamma(n - 1/2)$$

$$\frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 \frac{\Gamma(n+1/2)}{\Gamma(n-1/2)} = 2\sigma^2 (n - 1/2) \quad \left[\because \Gamma n = (n-1)\Gamma(n-1) \right]$$

(or) $\mu_{2n} = \sigma^2 (2n-1) \mu_{2n-2}$

which gives the recurrence relation for the moments of normal distribution

Additive Property of the Independent Normal Variates

Let X_1 and X_2 are two independent normal variates with mean μ_1, μ_2 and variance σ_1^2, σ_2^2

We know that $M_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}$

Now, $M_{X_1+X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(t)$

[$\because$ X_1 and X_2 are independent random variables]

$$= e^{\mu_1 t + \frac{t^2\sigma_1^2}{2}} \cdot e^{\mu_2 t + \frac{t^2\sigma_2^2}{2}}$$

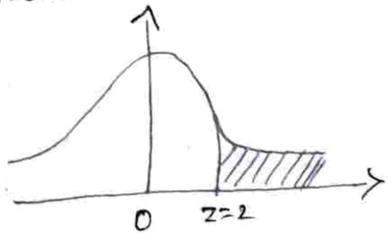
$$M_{X_1+X_2}(t) = e^{(\mu_1+\mu_2)t + \frac{t^2}{2}(\sigma_1^2+\sigma_2^2)}$$

which gives the m.g.f of the normal variate X_1+X_2 with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$. Hence the sum of independent normal variates is also a normal variate which gives the additive property of the normal distribution.

Examples:

1. X is normally distributed and the mean of X is 12 and the S.D is 4. Find out the probability of the following. (i) $x \geq 20$ (ii) $x \leq 20$ (iii) $0 \leq x \leq 12$.

Solution: Given $\mu = 12, \sigma = 4$.



(i) To find $p(x \geq 20)$

when $x = 20$, $z = \frac{x-\mu}{\sigma} = \frac{20-12}{4} = 2$.

(ii) when $x = 20, z = 2$

$\therefore p(x \geq 20) = p(z \geq 2) = 0.5 - p(0 \leq z \leq 2)$

$= 0.5 - 0.4772 = 0.0228$

(ii) To find $p(x \leq 20)$

when $x = 20$, $z = \frac{x-\mu}{\sigma} = \frac{20-12}{4} = 2$.

$\therefore p(x \leq 20) = p(z \leq 2) = 1 - p(z \geq 2) = 1 - 0.0228 = 0.9772$.

(iii) To find $p(0 \leq x \leq 12)$.

when $x = 0$, $z = \frac{x-\mu}{\sigma} = \frac{0-12}{4} = -3$

when $x = 12$, $z = \frac{x-\mu}{\sigma} = \frac{12-12}{4} = 0$

$\therefore p(0 \leq x \leq 12) = p(-3 \leq z \leq 0) = p(0 \leq z \leq 3) = 0.4987$.

2. The weekly wages of 1000 workmen are normally distributed around a mean of Rs. 70 with a S.D of Rs. 5. Estimate the number of workers whose weekly wages will be (i) between Rs. 69 and Rs. 72 (ii) less than Rs. 69 (iii) More than Rs. 72.

Soln: Given $\mu = 70$, $\sigma = 5$

(i) $P(69 < x < 72) = ?$

when $x = 69$, $z = \frac{x-\mu}{\sigma} = \frac{69-70}{5} = -0.2$

when $x = 72$, $z = \frac{x-\mu}{\sigma} = \frac{72-70}{5} = 0.4$

$\therefore P(69 < x < 72) = P(-0.2 < z < 0.4)$

$= P(-0.2 < z < 0) + P(0 < z < 0.4)$

$= P(0 < z < 0.2) + P(0 < z < 0.4)$

$= 0.0793 + 0.1554$

$= 0.2347$.

Out of 1000 workmen, the number of workers whose wages lies between Rs. 69 and Rs. 72 $= 1000 \times P(69 < x < 72)$

$= 1000 \times 0.2347 = 234.7 \simeq 235$

(ii) $P(\text{less than Rs. } 69) = P(x < 69) = ?$

when $x = 69$, $z = \frac{x-\mu}{\sigma} = \frac{69-70}{5} = -0.2$

$\therefore P(x < 69) = P(z < -0.2) = 0.5 - P(0 < z < 0.2)$

$= 0.5 - 0.0793 = 0.4207$.

out of 1000 workmen, the number of workers whose wages are less than Rs. 69 $= 1000 \times P(z < -0.2)$

$= 1000 \times 0.4207 = 420.7 \simeq 421$.

(iii) $P(\text{more than Rs. } 72) = P(x > 72)$

when $x = 72$, $z = \frac{x-\mu}{\sigma} = \frac{72-70}{5} = \frac{2}{5} = 0.4$

$\therefore P(x > 72) = P(z > 0.4) = 0.5 - P(0 < z < 0.4) = 0.5 - 0.1554$

$= 0.3446$.

Out of 1000 workmen, the number of workers whose wages are greater than 72 $= 1000 \times P(z > 0.4) = 1000 \times 0.3446 = 344.6 \simeq 345$