Probability and Statistics

Syllabus :
1. Introduction to probability
2. Additive & multiplicative laws of probability
3. Conditional probability
4. Independent events
5. Bayes theorem
6. Probability density function
7. Probability distribution function
8. Binomial, Poisson and Normal distribution
9. Mathematical expression of random variables
10. Moment generating function
11. Chebyshev's inequality
12. Central limit theorem

---

1. Curve fitting (fitting of straight line, parabola & exponential curves)
2. Correlation and Regression Analysis

---

1. 2-Dimensional Random Variables.

Book : Probability & statistics for Engineers by R.A. Johnson, I. Miller, J. Freund.

Random expt : An expt is said to be a random expt if the outcome maynot be predicted with certainity.

Sample space (S) : The set of collection of all possible outcomes of a random expt is called sample space.

eg. in a die rolling expt, $S = \{1, 2 \ldots 6\}$

Sample point / outcome / trial : Each element in a sample space is called a sample point.

ex. '4' is a sample point of the sample space in a die rolling expt.

ex. 'T' in a coin tossing expt.

Event (E) : Any subset of a sample space 'S' that favors to a particular performance is called an event.

ex. Consider a die rolling expt.

Let there be two events E_1, E_2 ∋

E_1 : Getting a no. $\leq 2 = \{1, 2\}$

E_2 : " " " $> 4 = \{5, 6\}$

Here E_1, E_2 are events.

Mutually Exclusive Events : Two events E_1 & E_2 ~~are~~ of same random expt are said to be mutually exclusive if $E_1 \cap E_2 = \phi$.

ex. E_1 & E_2 in the previous example are mutually exclusive.

Mutually Exhaustive Events : Two events E_1 & E_2 of the same random expt are said to be mutually exhaustive if (i) $E_1 \cap E_2 = \phi$

(ii) $E_1 \cup E_2 = S$

ex. in a die rolling expt

let E_1 : getting an odd number

E_2 : getting an even number

Here E_1 & E_2 are mutually exhaustive

ex. in a set of English alphabets

E_1 : vowels

E_2 : consonants

Equally likely events : Two events E_1 & E_2 are said to be equally likely if they have equal chances of occuring.

ex. getting H or T in a coin tossing expt.

OR, They have same probability.

Probability of an event : Let E be an event of a sample space S in a random expt. The probability of event E is defined as,

$$P(E) = \frac{\text{No. of favorable cases to E}}{\text{Total no. of ~~as~~ cases in S}}$$

$$\Rightarrow \quad P(E) = \frac{|E|}{|S|}$$

NOTE : $0 \leq P(E) \leq 1$

We know, $E \subseteq S$

$$\Rightarrow \quad |E| \leq |S|$$

$$\Rightarrow \quad 0 \leq |E| \leq |S|$$

$$\Rightarrow \quad 0 \leq \frac{|E|}{|S|} \leq \cancel{P(E)}\ 1$$

$$\therefore \quad 0 \leq P(E) \leq 1$$

26/7/16.

Independent events : Two events E_1 and E_2 are said to be independent if the occurence of one event doesn't depend on the occurence of the other.

In such case, we have,

$$\boxed{P(E_1 \cap E_2) = P(E_1)\,P(E_2)}$$

Conditional probability : Consider two events A and B, the conditional probability of an event 'A that B has already occured' is given by

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

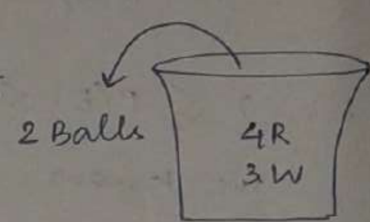
Similarly,
$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow \boxed{P(A \cap B) = P(A) \cdot P(B|A)}$$

<u>eg</u>: An urn contains 4 red and 3 white balls. Two balls are drawn at random one after another. Find the probability that the first ball drawn is red and the second ball drawn is white under the following conditions:

(i) with replacement.
(ii) without replacement.

Solⁿ: Define the following events



A : Getting a red ball
B : " " white ball

$\therefore \quad P(A) = \frac{4}{7}, \quad P(B) = \frac{3}{7}$

(i) <u>With replacement</u>, (independent events)

$$P(A \cap B) = P(A)\,P(B)$$
$$= \frac{4}{7} \cdot \frac{3}{7} = \frac{12}{49}$$

(ii) <u>Without replacement</u> (conditional probability)

$$P(A \cap B) = P(A)\,P(B|A)$$
$$= \frac{4}{7} \cdot \frac{3}{6} = \frac{12}{42} = \frac{2}{7}$$

Additive law of probability:

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

If A and B are mutually exclusive,

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

In general, for n events $E_1, E_2 \dots E_n$, we have

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i \neq j} P(E_i \cap E_j) + \sum_{i \neq j \neq k} P(E_i \cap E_j \cap E_k) - \dots + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n)$$

Eg: A pack of 52 playing cards is considered. A card is drawn at random from it. Find the probability that it is

(i) a red and an Ace card

(ii) a red or an Ace card.

Solⁿ: Define the events:

A : Getting a red card

B : " an ace card

$$\therefore P(A) = \frac{26}{52} = \frac{1}{2}, \quad P(B) = \frac{4}{52} = \frac{1}{13}$$

(i) $P(A \cap B) = P(A)\,P(B) = \frac{1}{2} \cdot \frac{1}{13} = \frac{1}{26}$

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

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

$$= \frac{15}{26} - \frac{1}{26} = \frac{14}{26} = \frac{7}{13}$$

eg: Two events cannot be both mutually exclusive and independent. Prove it.

(32)

Sol^n: consider two non zero events E_1 and E_2.

$P(E_1 \cap E_2) = P(E_1)\, P(E_2)$ [Independent events]

$P(E_1 \cap E_2) = \cancel{\phi}\ 0$ [Mutually exclusive]

But, $P(E_1)\, P(E_2) \neq \cancel{\phi}\ 0$ ($\because$ non zero events)

eg Prove that $P(A \cap B \cap C) = P(A)\, P(B|A) \cdot P(C|\overline{A \cap B})$

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

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

$= P(C|\overline{A \cap B}) \cdot P(A \cap B)$

$= P(C|\overline{A \cap B}) \cdot P(B \cap A)$

$= P(C|\overline{A \cap B}) \cdot P(B|A) \cdot P(A)$

02/08/16.

(Q1) What is the probability of obtaining atleast one head in tossing 8 fair coins?

Sol^n: ~~Let E: event such that~~

$$\boxed{\begin{aligned} P(E^c) &= 1 - P(E) \\ P(E) &= 1 - P(E^c) \end{aligned}}$$

P (Getting atleast one head)

$= 1 - P(\text{all } T)$

$= 1 - \frac{1}{2^8}$

$= \frac{2^8 - 1}{2^8} = \frac{255}{256}$

(Q2). 3 screws are drawn at random from a lot of 1000 screws, 10 of which are defectives. Find the probability of the event that all 3 screws drawn are non-defective, assuming that the screws are drawn

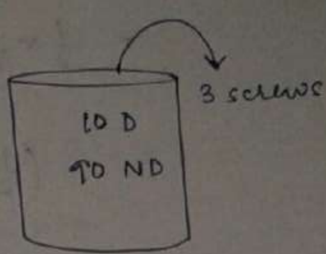
(i) with replacement.

(ii) without replacement.

Sol^n:

$$P(\text{getting a ND}) = \frac{90}{100} = \frac{9}{10}$$

$$P(\text{getting a D}) = \frac{10}{100} = \frac{1}{10}.$$



$$\cancel{r - \frac{9}{10} \cdot \frac{9}{10}}$$

(i)

E_1 : 1st drawn is ND

E_2 : 2nd drawn is ND

E_3 : 3rd drawn is ND

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)\,P(E_2)\,P(E_3)$$

$$= \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{9}{10} = (0.9)^3$$

(ii) $$P(E_1 \cap E_2 \cap E_3) = P(E_1)\,P(E_2|E_1)\,P(E_3|\overline{E_1 \cap E_2})$$

$$= \frac{90}{100} \times \frac{89}{99} \times \frac{88}{98}$$

(Q3) Suppose we draw cards repeatedly and with replacement from a lot of 200 cards, 100 of which refers to 'MALE' and 100 to 'FEMALE' persons. What is the probability of obtaining the third 'MALE' card before the second 'FEMALE' card? ~~Wri~~ Moreover, write the complement of this question and answer it.

Sol^n:

$$P(\text{Male}) = \frac{100}{200}, \quad P(\text{Female}) = \frac{100}{200}$$

$$\cancel{\frac{1}{100} \times \frac{1}{100}}$$

$$\cancel{P(M|F) = \frac{1}{197}}$$

P

P

$P(3^{rd}\, M < 2nd\, F)$

$= P(MMM) + P(FMMM) + P(MFMM) + P(MMFM)$

$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \cancel{0}\frac{1}{2} \times 3$

$= \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 3 = \frac{5}{16}.$

Complement question : $P(2^{nd}\, F < 3^{rd}\, M)$.

Probability $= 1 - P(3^{rd}\, M < 2^{nd}\, F)$

$= 1 - \frac{5}{16}$

$= \frac{11}{16}.$

---

(84) If $P(A) > P(B)$, then prove that

$$P(A|B) > P(B|A)$$

Sol^n:

$$P(A) > P(B)$$

$$\cancel{P(A \cap B)\, P(A) > P(A \cap B)\, P(B)}$$

$$\frac{1}{P(A)} < \frac{1}{P(B)}$$

$$\frac{P(A \cap B)}{P(A)} < \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) < P(A|B)$$

(Q5) ~~What abstract~~ If two events A and B are independent, then show that A^c and B^c are also independent.

Sol^n:

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

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

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

$$\Rightarrow P(A\cup B)^c = 1 - P(A) - P(B) + P(A).P(B).$$

$$\therefore P(A^c\cap B^c) = \cancel{1 - P(A)[P(B) - 1] - P(B)}$$

$$\cancel{P(A)(1+P(B))}$$

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

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

$$= P(A^c)\,P(B^c)$$

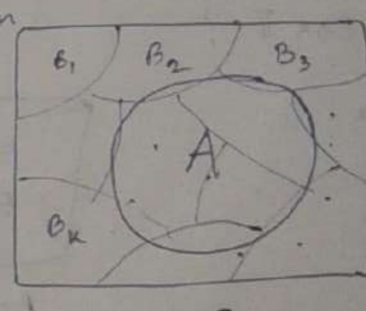
__Baye's theorem:__

Let $B_1, B_2 \dots B_k$ be a partition of the sample space S. Let A be an event associated with S.

Then,
$$P(B_i|A) = \frac{P(A|B_i).P(B_i)}{\sum_{j=1}^{k} P(A|B_j)\,P(B_j)}, \quad i = 1, 2 \dots k$$

09/08/16

__Proof:__ Given, $B_1, B_2 \dots B_k$ is a partition of the sample space S.



S

$$\Rightarrow B_i\cap B_j = \phi \quad \forall\ i\neq j$$

$$\& \bigcup_{j=1}^{k} B_j = S$$

$$(A\cap B_1)\cup(A\cap B_2) \dots \cup(A\cap B_k)$$

$$= \{A\cap(B_1\cup B_2)\}\cup(A\cap B_3) \dots \cup(A\cap B_k)$$

$$= \dots$$

$$= A\cap(B_1\cup B_2\cup \dots \cup B_k)$$

$$= A\cap S = A \quad [\because A\subseteq S]$$

$$\longrightarrow (1)$$

Now $\quad P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_k)$

$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots P(A|B_k)$$

$$= \sum_{j=1}^{k} P(A|B_j)\,P(B_j) \quad \text{——} (2)$$

$$LHS = P(B_i|A) = \frac{P(A \cap B_i)}{P(A)}$$

$$= \frac{P(A|B_i)\,P(B_i)}{\sum_{j=1}^{k} P(A|B_j)\,P(B_j)}$$

$$= RHS.$$

Hence, proved.

(Q) Consider there are 3 boxes containing red balls and black balls as follows:

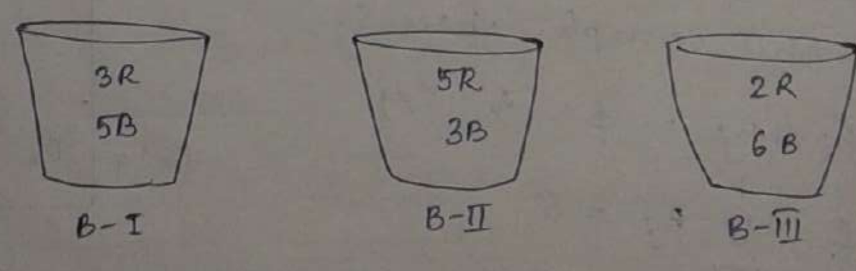
Box I : 3R and 5B

Box II : 5R and 3B

Box III : 2R and 6B

One ball is drawn at random from an arbitrarily selected box. If ball comes out to be black, then what is the probability that it is drawn from Box II ?

Sol^n:



Let us define the following events,

A : Getting a black ball

B_i : Selecting the i^{th} box

We need to find

$$P(B_2|A) = \frac{P(A|B_2)\,P(B_2)}{\sum_{j=1}^{3} P(A|B_j)\,P(B_j)} \quad \text{——} (i)$$

$P(A|B_2) = \frac{3}{8}$

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

$P(A|B_1) = \frac{5}{8}$

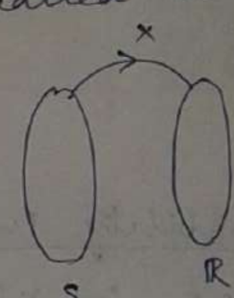
$P(A|B_3) = \frac{6}{8}$

$$\therefore \quad P(B_2|A) = \frac{\frac{3}{8}\cdot\frac{1}{3}}{\frac{5}{8}\cdot\frac{1}{3} + \frac{3}{8}\cdot\frac{1}{3} + \frac{6}{8}\cdot\frac{1}{3}}$$

$$= \frac{3}{5+3+6} = \frac{3}{14}$$

Random Variable :-

A real-valued function defined on the sample space S, of a random experiment is called a random variable.

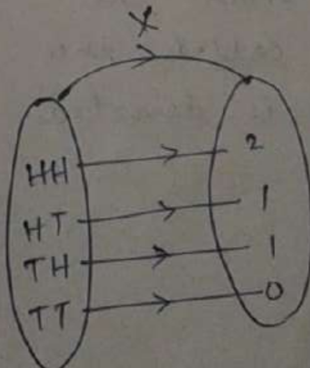


eg. Two coins are tossed once.

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

X : Number of H's.

Let X be the R.V.
is defined as X : No. of heads
Then X is picterized as



at) P.F. $\rightarrow f(X=x) = f(x)$

P.D $\rightarrow (X, f(x))$

ufto) D.F. $\rightarrow F(x)$

Random Variable / Stochastic Variable

Discrete R.V.

1. **Defn**

A R.V. that assumes a countable no. of real values is called a discrete R.V.

2. **Ex:**

No. of print mistakes in a book, pagewise.

~~No. of~~ Marks obtained by students of a particular class.

3. **Probability Function $f(x)$**

~~The P.F~~ Let X be a R.V. The p.f. is defined as

$$f(x) = P(X = x)$$
$$= \begin{cases} p_j, & P(X = x_j) \\ 0, & \text{otherwise} \end{cases}$$

4. **Probability distribution**

A R.V. X along with the prob. f^n $f(x)$ is called the prob. dist^n, and is denoted by $(X, f(x))$.

5. **Properties**

(i) The prob. function is also known as prob. mass function (PMF)

Continuous R.V.

1. **Defn**

A R.V. that assumes all possible real values within a given range is called a continuous R.V.

2. **Ex:**

Heights of college students in India.

The rainfall rate in a particular time interval.

Temperature reading in a room over time.

3. **Probability Function**

Let X be a cont. R.V. The p.f. is defined as

$$f(x) = F'(x)$$

where, $F(x)$ is the distribut^n function.

4. **Probability distribution**

— DO —

5. **Properties**

(i) The prob. function is also known as prob. density function (PDF).

(ii) $0 \leq p_j \leq 1$

(iii) $\sum_j p_j = 1$

6. <u>Distribution function</u>

$$F(x) = P(X \leq x)$$
$$= \sum_{x_j \leq x} P(X = x_j)$$

7. <u>Ex</u>:

Let a die is rolled once.

$X = x$	1	2	3	4	5	6
AT $f(x)$ $P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6
$F(x)$ $P(X \leq x)$	1/6	2/6	3/6	4/6	5/6	1

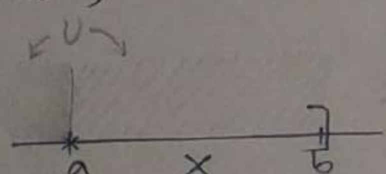
UPTO

$F(4)$
$= P(X \leq 4)$

<u>Properties</u>

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

<u>Pf</u>



$(X \leq b) = (X \leq a) \cup (a < X \leq b)$

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

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

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

(ii) $f(x)$ is integrable over $(-\infty, \infty)$

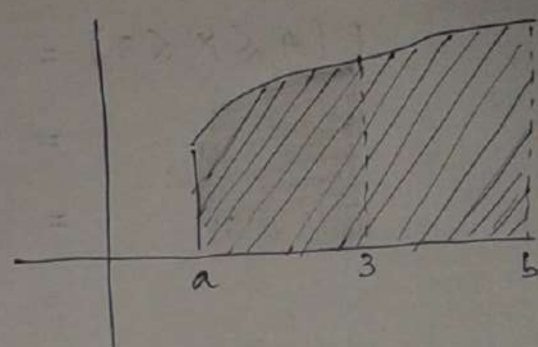
(iii) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

6. <u>Distribution function</u>

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

or cdf (cumulative distrib. f^n).

7. <u>Ex</u>:



$\int_a^b f(x)\, dx = \text{Area} = 1$

$F(3) = \text{Area upto } 3.$

$f(3) = \text{Area at } 3 = 0.$

<u>Properties</u>

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

→ inclusion of a point does not make any difference.

Ex 1: Two dice are rolled once. Find the probability of getting the sum of both the outputs atleast 4 and atmost 8.

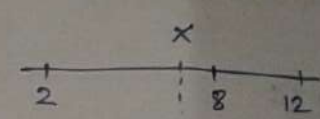
Sol^n: Let us defined a R.V. X such that

X : sum of both outcomes

$$\therefore \cancel{P(4 \leq X \leq 8)} = \cancel{P(4) + P(4 < X \leq 8)}$$

$$= \cancel{P(4) + F(8) - F(4)}.$$

$$\cancel{P(4) = \frac{\quad}{36}}$$



$$P(4 \leq X \leq 8) = P(3 < X \leq 8)$$

$$= F(8) - F(3)$$

$$= P(X \leq 8) - P(X \leq 3)$$

$$= [1 - P(X > 8)] - P(X \leq 3)$$

1 1	2 1	3 1	4 1	5 1	6 1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6

$$= \left(1 - \frac{10}{36}\right) - \frac{3}{36}$$

$$= \frac{23}{36}$$

Ex 2: Let a coin is tossed repeatedly. Let X be a R.V. of trials upto the first H appears. Show that

$$\sum_j P_j = 1.$$

Sol^n: Define $P_j = P(X = x_j)$

where x_j : getting H on j^{th} trial.

Now, $\sum_j x_j = x_1 + x_2 + x_3 + \cdots$

$= P(H) + P(TH) + P(TTH) + \cdots$

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

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

Ex3: Check whether the function $f(x) = 4x^3$, $0 \leq x \leq 1$ is a pdf?

Sol^n: check $\int_{-\infty}^{\infty} f(x)\,dx = 1$ i.e. $\int_0^1 f(x)\,dx = 1$.

$$\int_0^1 4x^3\,dx = x^4 \Big]_0^1$$

$$= 1.$$

$\therefore$ The function is a pdf.

Ex4: The PMF of a R.V. is given by $P(i) = \frac{c\lambda^i}{i!}$, for $i = 0, 1, 2 \ldots$

λ being a positive value.

Find (i) $P(X=0)$ (ii) $P(X>3)$

Sol^n: $\sum_{i=0}^{\infty} P(i) = 1$

$\Rightarrow P(0) + P(1) + P(2) + P(3) + \cdots = 1$

$\Rightarrow c + c\lambda + \frac{c\lambda^2}{2!} + \frac{c\lambda^3}{3!} + \cdots = 1$

$\Rightarrow c\left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right] = 1$

$\Rightarrow c\,e^\lambda = 1$

$\Rightarrow c\,(e^\lambda) = 1$

$\Rightarrow e^\lambda = \frac{1}{c}$

$\Rightarrow e^\lambda = \frac{1}{c}$

$\Rightarrow \boxed{c = e^{-\lambda}}$

$$P(i) = \frac{e^{-\lambda} \cdot \lambda^i}{i!}$$

(i) $P(x=0) = P(0) = e^{-\lambda}$

(ii) $P(x>3) = 1 - P(x \leq 3)$

$$= 1 - \{ P(x=0) + P(x=1) + P(x=2) + P(x=3) \}$$

$$= 1 - \left\{ e^{-\lambda} + \frac{\lambda e^{-\lambda}}{1!} + \frac{\lambda^2}{2!} e^{-\lambda} + \frac{\lambda^3}{3!} e^{-\lambda} \right\}$$

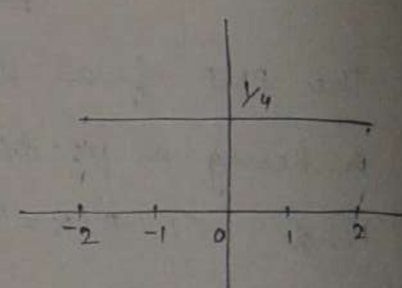
$$= 1 - e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right)$$

8 A R.V. X has the Probability density f^n (PDF)

$$f(x) = \begin{cases} \frac{1}{4}, & -2 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Find (i) $P(X<1)$ (ii) $P(|x|>1)$

(iii) $P(2x+3>5)$



Solⁿ: (i) $P(X<1) = F(1)$

$$= \int_{-\infty}^{1} \frac{1}{4}\, dt$$

$$= \int_{-2}^{1} \frac{1}{4}\, dt$$

$$= \frac{1}{4}[t]_{-2}^{1} = \frac{3}{4}$$

$|x|>1$

$1 - (-1 \leq x \leq 1)$

(ii) $P(|x|>1) = 1 - \int_{-1}^{1} \frac{1}{4}\, dt$

$$= 1 - \frac{1}{4}[t]_{-1}^{1} = 1 - \frac{1}{4}[1+1] = \frac{1}{2}$$

(iii) $P(2x+3>5) = P(x>1)$

$$= \int_{1}^{2} \frac{1}{4}\, dt = \frac{1}{4}$$

⑧ Let X be a continuous R.V. having the PDF

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

Find the value of k and the cdf $\rightarrow F(x)$

Solⁿ: (i)

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

$$\Rightarrow \int_0^1 k(1-x^2)\, dx = 1$$

$$\Rightarrow k \int_0^1 1-x^2\, dx = 1$$

$$\Rightarrow k \left[x - \frac{x^3}{3} \Big|_0^1 \right] = 1$$

$$\Rightarrow k \left[1 - \frac{1}{3} \right] = 1$$

$$\Rightarrow k = \frac{3}{2}$$

(ii)

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

$$= \int_0^x \frac{3}{2}(t^2-1)\, dt$$

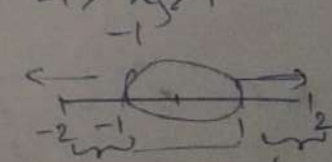
$$= \frac{3}{2}\left[\frac{t^3}{3} - t \right]_0^x$$

$$= \frac{3}{2}\left(\frac{x^3}{3} - x \right) = \frac{x^3}{2} - \frac{3x}{2}$$

$$= \frac{x}{2}(x^2-3)$$

-2 2

$-1 > x \geq 1$



Mean and Variance of a probability Distribution :-

Mean :-
$$\mu = \sum x_j f(x_j) \quad [\text{ for discrete R.V. } x]$$

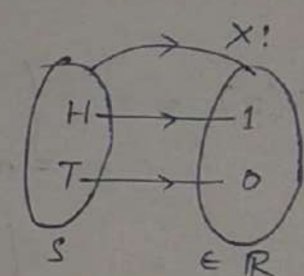
OR
$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx \quad [\text{ for continuous R.V. } x]$$

Variance :-
$$\sigma^2 = \sum_j (x_j - \mu)^2 f(x_j) \quad [\text{ for discrete}]$$

OR
$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \quad [\text{ for continuous}]$$

Ex1. A coin is tossed once. Find the mean and variance of the probability distribution by considering the R.V. X as the number of heads.

Sol^n.
$$S = \{ H, T \}$$



	H	T
$X = x$	x_1 1	x_2 0
$f(x) = P(X=x)$	$f(x_1)$ $\frac{1}{2}$	$f(x_2)$ $\frac{1}{2}$

$$\therefore \mu = \sum_{j=1}^{2} x_j f(x_j) = x_1 f(x_1) + x_2 f(x_2)$$
$$= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 = \frac{1}{2}[1+0]$$
$$= \frac{1}{2}$$

And,
$$\sigma^2 = \sum_{j=1}^{2} \left(x_j - \frac{1}{2}\right)^2 f(x_j)$$
$$= \left(x_1 - \frac{1}{2}\right)^2 f(x_1) + \left(x_2 - \frac{1}{2}\right)^2 f(x_2)$$
$$= \frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{4}$$

Ex 2. Two coins are tossed once, and X be the R.V. of getting number of heads.

Sol^n: $S = \{HH, HT, TH, TT\}$

	HH	TT	HT	TH
$X = x$	x_1 2	x_2 0	x_3 1	x_4 1
$f(x) = P(X=x)$	$f(x_1)$ $1/4$	$f(x_2)$ $1/4$	$f(x_3)$ $1/4$	$f(x_4)$ $1/4$

$$\therefore \mu = \sum_{j=1}^{4} x_j f(x_j) = \left(2\times\frac{1}{4} + 0\times\frac{1}{4}\right) + \left(1\times\frac{1}{4} + 1\times\frac{1}{4}\right)$$

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

And $\sigma^2 = \sum_{j=1}^{4} (x_j - 1)^2 f(x_j)$

$$\cancel{= \left(2-\frac{3}{2}\right)^2 \times \frac{1}{4} + \left(0-\frac{3}{2}\right)^2 \times \frac{1}{4}}$$

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

$$\cancel{= \frac{10}{16} + \frac{1}{4} = \frac{5+2}{8} = \frac{7}{8}}$$

$$= \left\{(2-1)^2 + (0-1)^2 + (1-1)^2 + (1-1)^2\right\} \frac{1}{4}$$

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

30/8/16

Expectation :

Let X be a R.V. with probability function $f(x)$.
The expectation of X is defined as

$$E(X) = \text{Mean} = \mu = \sum_x x\, f(x) \quad (\text{for } X \text{ Discrete})$$

$$\text{OR} \quad = \int_{-\infty}^{\infty} x\, f(x)\, dx \quad (\text{for } X \text{ Continuous})$$

In general,

$$E[g(X)] = \begin{cases} \sum_x g(x) f(x) \ ; & \text{Discrete } X \\ \int_{-\infty}^{\infty} g(x) f(x)\, dx \ ; & \text{Continuous} \end{cases}$$

Properties

1. $E(a) = a$, 'a' being a constant
2. $E(aX) = a\,E(X)$, a being a constant
3. $E(X \pm Y) = E(X) \pm E(Y)$
4. $E(XY) = E(X) \cdot E(Y)$, provided X and Y are independent

NOTE : $\boxed{\text{Variance} = \sigma^2 = E(X^2) - \{E(X)\}^2}$

Pf :-

$$\text{Variance} = \sigma^2 = E(X-\mu)^2$$

$$\sigma^2 = \sum_x \underbrace{(x-\mu)^2}_{g(x)} f(x) = E(x-\mu)^2$$

$$= E(X^2 - 2\mu X + \mu^2)$$

$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$

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

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

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

Proved.

Q1. Let a die is rolled once and X be the R.V. that appears in the roll. Find

i) $E(X)$ and ii) $E(2X+1)$

Sol^n:

S	1	2	3	4	5	6
$X=x$	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

i) $E(X) = \sum_{x=1}^{6} x.f(x)$

$= \frac{1}{6}[1+2+\dots 6]$

$= \frac{1}{\cancel{6}} \cdot \frac{\cancel{6}.7}{2} = \frac{7}{2}$

ii) $E(2X+1) = \sum_{x=1}^{6} (2x+1) f(x)$

$= \frac{1}{6}[3+5+7+9+11+13]$

$= \frac{1}{6} \cdot 48 = 8$.

or, $E(2X+1)$

$= 2E(X)+1$

$= 7+1 = 8$.

Q2. The PDF of a R.V. X is

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

Find (i) $E(X)$ and (ii) $E(2X^3)$

Sol^n: $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-1}^{1} x \frac{1}{2}\, dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_{-1}^{1}$

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

$= 0$

$E(2X^3) = 2E(X^3)$

$= 2\int_{-\infty}^{\infty} x^3 f(x)\, dx = \int_{-1}^{1} x^3\, dx = \frac{x^4}{4}\Big|_{-1}^{1}$

$= 0$.

Q3. A random number is chosen from a set $\{1, 2, \ldots 100\}$ and another from $\{1, 2, \cdots, 50\}$. What is the expectation of their product?

Sol^n: Let us define the R.V.

X : selecting a number from $\{1, 2, \ldots 100\}$

Y : " " " " $\{1, 2, \ldots 50\}$

$$E(XY) = E(X)\,E(Y) \quad \text{——} \ (1)$$

$$E(X) = \frac{1}{100}\{1+2+\ldots 100\} = \frac{1}{100} \cdot \frac{100 \cdot 101}{2} = \frac{101}{2}$$

$$E(Y) = \frac{1}{50}\{1+2+\ldots 50\} = \frac{1}{50} \cdot \frac{50 \cdot 51}{2} = \frac{51}{2}.$$

$$\begin{array}{r} 101 \\ \times 51 \\ \hline 101 \\ 505 \\ \hline 5151 \end{array}$$

$$\therefore (1) \Rightarrow E(XY) = \frac{101 \times 51}{2 \times 2} = \frac{5151}{4}$$

Q4 A special die with $(n+1)$ faces is rolled. Its faces are marked $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}$.

If X denotes the number shown, then find

(i) $E(X)$ (ii) S.D. of X (iii) $E\left(X - \frac{1}{2}\right)^3$

Sol^n: (i) $E(X) = \dfrac{1}{n+1}\left\{\dfrac{1+2+\ldots n-1+n}{n}\right\}$

$= \dfrac{1}{2}$

(ii) $\text{S.D.} = \sigma = \sqrt{E(X^2) - \{E(X)\}^2}$

$= \sqrt{\frac{1}{2} \cdot \frac{1}{2} - \left(\frac{1}{2}\right)^2} = 0$

(iii) $E\left(X - \frac{1}{2}\right)^3 = E\left(X^3 + \frac{1}{8} - \frac{3X^2}{2} + \frac{3X}{4}\right)$

$= E(X^3) + \frac{1}{8} - \frac{3}{2}E(X^2) + \frac{3}{4}E(X)$

$$E(x^2) = \frac{2n+1}{6n}$$

(ii) $$E(x^2) = \frac{1}{n+1} \left\{ \frac{1^2+2^2+\dots (n-1)^2+n^2}{n^2} \right\}$$

$$= \frac{1}{n+1} \cdot \frac{n(n+1)(2n+1)}{6n^2}$$

$$= \frac{2n+1}{6n}$$

$$\therefore \quad S.D. = \sigma = \sqrt{\frac{2n+1}{6n} - \frac{1}{4}}$$

$$= \sqrt{\frac{4n+2-3n}{12n}} = \sqrt{\frac{n+2}{12n}}$$

(iii) $$E\left(x-\frac{1}{2}\right)^3 = E(x^3) + \frac{1}{8} - \frac{3}{2} E(x^2) + \frac{3}{4} E(x)$$

$$= \frac{3}{2} \frac{2n+1}{6n} + \frac{1}{8} + \frac{3}{4} \frac{1}{2}$$

Q5. A RV X has pdf $f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$

Find i) $E(X)$ ii) $E(X^2)$

iii) $E(X-1)^2$ iv) $E\left(e^{2X/3}\right)$

Soln:

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

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

$$= \left[x \frac{e^{-x}}{-1}\right]_0^{\infty} - \int_0^{\infty} \frac{e^{-x}}{-1}\,dx$$

$$= 0[0-1] - e^{-x}\Big|_0^{\infty}$$

$$= -[0-1] = 1.$$

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

$$= x^2 e^{-x}\Big|_0^{\infty} + 2\int_0^{\infty} x e^{-x}\,dx$$

$$= 2$$

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

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

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

$$= 2 + 1 - 2 = 1.$$

(iv) $$E\left(e^{2X/3}\right) = \int_0^{\infty} e^{2x/3} e^{-x}\,dx$$

$$= \int_0^{\infty} e^{-x/3}\,dx = -3e^{-x/3}\Big|_0^{\infty}$$

$$= -3(0-1) = 3.$$

Alternatively,

(i) $E(X) = \int_0^\infty x e^{-x} dx = \Gamma(2)$
$= 1! = 1.$

(ii) $E(X^2) = \int_0^\infty x^2 e^{-x} dx = \Gamma(3) = 2! = 2$

06/09/16.

Moments and Moment Generating Function :

Moment (i) r^{th} order moment about the origin :-

$$\mu_r' = E(X^r) = \begin{cases} \sum_x x^r \cdot f(x) & ; X : \text{Discrete} \\ \int_{-\infty}^{\infty} x^r f(x)\, dx & ; X : \text{continuous} \end{cases}$$

(ii) r^{th} order moment about any point 'c' :-

$$E(X-c)^r = \begin{cases} \sum_x (x-c)^r f(x) & ; X : \text{Discrete} \\ \int_{-\infty}^{\infty} (x-c)^r f(x)\, dx & ; X : \text{Cont.} \end{cases}$$

(iii) r^{th} order moment about Mean (or)
r^{th} order central moment

$$M_r' = E(X-\mu)^r = \begin{cases} \sum_x (x-\mu)^r f(x) & ; X : \text{Discre} \\ \int_{-\infty}^{\infty} (x-\mu)^r f(x)\, dx & ; X : \text{Cont} \end{cases}$$

NOTE 1^{st} order moment about origin = Mean
$= E(X)$
$= \mu$

Q Find the 1st and 2nd order central moment

Soln: $M_1' = E(X-\mu)$

$= E(X) - E(\mu)$

$= \mu - \mu = 0$

NOTE: 1st order CM for any data $= 0$.

$M_2' = E(X-\mu)^2$

$= E(X^2) - 2\mu E(X) + E(\mu^2)$

$= E(X^2) - 2\mu^2 + \mu^2$

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

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

$= \sigma^2 =$ Variance

NOTE: 2nd order CM for any data = Variance of the data.

Moment Generating Function (MGF):

Let X be a RV with prob. function $f(x)$. The MGF of X about origin is given by $M_X(t)$ and is defined by

$$M_X(t) = E(e^{tx}) = \begin{cases} \sum_x e^{tx} f(x) \ ; & X: \text{Discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \ ; & X: \text{cont.} \end{cases}$$

Thm:- Let X be a RV with prob. function $f(x)$ and MGF $M_X(t)$. Then the r^{th} order moment about the origin is given by

$$\boxed{\mu_r' = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0}}$$

Proof :

RHS $$\frac{d^r}{dt^r} M_X(t) = \frac{d^r}{dt^r} E(e^{tx})$$

$$= \begin{cases} \sum_x x^r e^{tx} f(x) & ; \text{Disc } X \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)\, dx & ; \text{Cont. } X. \end{cases}$$

Putting $t = 0$,

$$\left.\frac{d^r}{dt^r} M_X(t)\right|_{t=0} = \begin{cases} \sum x^r f(x) & ; \text{Disc } X \\ \int_{-\infty}^{\infty} x^r f(x)\, dx & ; \text{Cont. } X \end{cases}$$

$$= E(X^r) = \mu_r'$$

= r^{th} order moment about the origin = LHS.

Binomial Distribution :-

[Bernoulli's Experiment : An expt. is called Bernoulli's expt if each of its trial (called Bernoulli's trial) has equal chances of occurring (or) they have equal probabilities.]

Let there be a Bernoulli's expt in which

p : probability of success

q : " " failure ($q = 1-p$)

n : the total no of Bernoulli's trial in the exp

x : the total no of success positions out achieved out of n trials.

Rearranging the structure by keeping all success positions initially, it will look like

$$\underbrace{\boxed{S}\ \boxed{S}\ \boxed{S}\ \cdots\ \boxed{S}}_{x \text{ times}}\ \underbrace{\boxed{F}\ \cdots\ \boxed{F}\ \boxed{F}}_{(n-x) \text{ times}}$$

The probability of the above structure $= p^x q^{n-x}$

However, the x success positions out of n trials can be chosen in ${}^nC_x$ ways. Therefore,

total probability of getting x success positions out of n trials is given by,

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

Now, the discrete R.V. X along the pmf $f(x)$ defined above together is called Binomial Distribution.

Q. Let a coin is tossed 3 times. Find the prob. of getting exactly two heads by using binomial distribution.

Sol^n

$$P(X=2) = {}^3C_2 \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^{3-2}$$

$$= {}^3C_2 \; \tfrac{1}{4} \cdot \tfrac{1}{2}$$

$$= \frac{3!}{2!} \cdot \frac{1}{8} = \frac{3}{8}$$

20/9/16.

Q Find the probability of getting atleast 2 six when a die is rolled four times.

Sol^n

$$\cancel{P(X=2) + P(X=3) + P(X=4) = {}^4C_2 \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^2 + = \frac{4!}{2!} \quad \frac{25}{1296}}$$

Let X : No. of 'six'

$n = 4$

$x = 2, 3, 4$

$p = \frac{1}{6}$

$q = \frac{5}{6}$

$$P(x>2) = 1 - P(x<2)$$
$$= 1 - \{ P(x=0) + P(x=1) \}$$
$$= 1 - \{ {}^4C_0 \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^4 + {}^4C_1 \left(\tfrac{1}{6}\right)^1 \left(\tfrac{5}{6}\right)^3 \}$$
$$= 1 - \{ \frac{5^4}{6^4} + 4 \times \frac{5^3}{6^4} \}$$
$$= 1 - \frac{9 \times 5^3}{6^4} = \frac{6^4 - 9 \times 125}{6^4}$$
$$= \frac{19}{144}.$$

Q Let a coin is tossed three times, what is the probability of getting atleast one tail?

Sol^n. Let x : Getting a tail.

$n = 3$

$X = 1, 2, 3$

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

$q = \frac{1}{2}$

$$P(x>1) = 1 - P(x<1)$$
$$= 1 - \{ P(x=0) \}$$
$$= 1 - {}^3C_0 \left(\tfrac{1}{2}\right)^0 \left(\tfrac{1}{2}\right)^3$$
$$= 1 - \frac{1}{8} = \frac{7}{8}.$$

Q Find the mean and variance for the Binomial distribution.

Sol^n. Mean $= \mu = E(X) = \sum_{x=0}^{n} x f(x)$

$$= \sum_{x=1}^{n} x f(x)$$
$$= \sum_{x=1}^{n} x \, {}^nC_x \, p^x q^{n-x}$$
$$= {}^nC_1 \, p \, q^{n-1} + 2 \cdot {}^nC_2 \, p^2 q^{n-2} + \dots$$
$$\dots n \, {}^nC_n \, p^n q$$

$$= npq^{n-1} + 2\,{}^nC_2\, p^2 q^{n-2} + \dots + n\cdot p^n$$ 8.

$$= np\left[q^{n-1} + 2\,\frac{{}^nC_2}{n} p q^{n-2} + \dots p^{n-1} \right]$$

$$= np\,[q+p]^{n-1}$$

$$= np$$

$$\therefore \boxed{\text{Mean} = \mu = E(X) = np}$$

<u>Variance</u> $= \sigma^2 = E(X^2) - [E(X)]^2 \longrightarrow ①$

Now, $E(X^2) = \sum_{x=0}^{n} x^2 f(x)$

$$= \sum_{x=0}^{n} \{x + (x-1)x\} f(x)$$

$$= \sum_{x=0}^{n} x f(x) + \sum_{x=0}^{n} (x-1)x f(x)$$

$$= np + \sum_{x=2}^{n} (x-1)\, x\, {}^nC_x\, p^x q^{n-x}$$

$$= np + 1\cdot 2\,{}^nC_2\, p^2 q^{n-2} + 2\cdot 3\,{}^nC_3\, p^3 q^{n-3} + \dots (n-1)\cdot n\,{}^nC_n\, p^n$$

$$= np + n(n-1)p^2\left[q^{n-2} + \frac{2\cdot 3\,{}^nC_3}{n(n-1)} p \dots + \dots p^{n-2} \right]$$

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

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

By eq^n ①, $\sigma^2 = np + n(n-1)p^2 - [np]^2$

$$= np + \cancel{n^2p^2} - np^2 - \cancel{n^2p^2}$$

$$= np(1-p) = npq.$$

$$\boxed{\text{Variance} = \sigma^2 = npq.}$$

8. Find the mean and variance for a binomial distribution by using MGF.

Mean $= \mu = E(X) = \frac{d M_x(t)}{dt}\Big|_{t=0}$

$$M_x(t) = E(e^{tx}) = \sum_{x=0}^{n} e^{tx} f(x)$$

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

$$= {}^nC_0\, p^0 q^n + e^t\, {}^nC_1\, p\, q^{n-1} + e^{2t}\, {}^nC_2\, p^2 q^{n-2} + \dots \; e^{tn}\, {}^nC_n\, p^n$$

$$= (pe^t)^0 q^n + {}^nC_1 (pe^t)^1 q^{n-1} + \dots \dots + (pe^t)^n$$

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

$$\therefore \mu = \frac{d}{dt}(q+pe^t)^n \Big|_{t=0}$$

$$= \left[n(q+pe^t)^{n-1} pe^t \right]_{t=0}$$

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

$$= \underline{\underline{np}}.$$

Variance $= \sigma^2 = E(X^2) - [E(X)]^2$

$$E(X^2) = \frac{d^2}{dt^2}(q+e^t p)^n \Big|_{t=0}$$

$$= \frac{d}{dt}\, n(q+pe^t)^{n-1} pe^t$$

$$= n(n-1)(q+pe^t)^{n-2} p^2 e^{2t} + n(q+pe^t)^{n-1} pe^t$$

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

$$\therefore \sigma^2 = \cancel{n^2p^2} - np^2 + \cancel{\;} - \cancel{n^2p^2} + np$$

$$= \underline{\underline{npq}}.$$

Poisson's Distribution

04/10/16

(5)

Let X be a RV (discrete RV) with prob. mass function $f(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$; $\lambda = np$

Now, X along p with f(x) defined above is called a Poisson distribution.

Usefulness

(i) If $n \to \infty$
(ii) If $p \to 0$ $\Rightarrow$ Poisson distribution

Theorem :- Poisson distribⁿ is a limiting case of Binomial distribⁿ.

Proof By B.D, the pmf is

$$f(x) = {}^{n}C_{x}\, p^{x} q^{n-x} \quad ; x = 0, 1 \ldots n$$

$$= \frac{n!}{x!(n-x)!} p^{x} q^{n-x} \quad ; x = 0, 1 \ldots n$$

$\because \lambda = np$

$p = \frac{\lambda}{n}$

$q = 1 - p$

$= 1 - \frac{\lambda}{n}$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

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

Put $n \to \infty$

$$= \frac{\lambda^{x}}{x!} e^{-\lambda} \cdot 1.$$

$$= \frac{e^{-\lambda}\lambda^{x}}{x!} \quad ; x = 0, 1 \ldots \infty$$

$\therefore$ PD is a limiting case of B.D.

Find the mean and variance for Poisson distribution.

Mean $= \mu = E(X) = \sum_{x=0}^{\infty} x f(x)$

$$= \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{(x-1)!}$$

$$= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{(x-1)!}$$

$$= e^{-\lambda}\left[\lambda + \frac{\lambda^2}{1!} + \frac{\lambda^3}{2!} + \frac{\lambda^4}{3!} + \cdots\right]$$

$$= \lambda e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right]$$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$\boxed{\text{Mean} = \lambda = np}$$

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

$$\therefore \quad E(X^2) = \sum_{x=0}^{\infty} x^2 f(x)$$

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

$$= \sum_{x=1}^{\infty} \frac{x\, e^{-\lambda}\lambda^{x}}{(x-1)!}$$

$$= e^{-\lambda}\left[\lambda + \frac{2\lambda^2}{1} + \frac{3\lambda^3}{2!} + \frac{4\lambda^4}{3!} + \cdots\right]$$

$$= e^{-\lambda}\lambda\left[1 + \frac{2\lambda}{1!} + \frac{3\lambda^2}{2!} + \frac{4\lambda^3}{3!} + \cdots\right]$$

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

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

$$= \lambda + \sum_{x=2}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{(x-2)!}$$

$$= \lambda + e^{-\lambda}\left[\lambda^2 + \frac{\lambda^3}{1!} + \frac{\lambda^4}{2!} + \frac{\lambda^5}{3!} + \cdots\right]$$

$$= \lambda + e^{-\lambda}\lambda^2\left[e^{\lambda}\right] = \lambda + \lambda^2$$

$$=$$

$$\therefore \sigma^2 = E(X^2) - \{E(X)\}^2$$
$$= \lambda + \lambda^2 - \lambda^2$$
$$= \lambda = np$$

Q1 For a Poisson variate, if $P(X=2) = P(X=1)$ then find $P(X=1$ or $X=0)$
Also, find the mean of the distribution.

Soln: Given, $P(X=2) = P(X=1)$

$$\Rightarrow f(2) = f(1)$$
$$\Rightarrow \frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-\lambda}\lambda}{1!}$$
$$\Rightarrow \frac{\lambda^2}{2} = \lambda$$
$$\Rightarrow \lambda^2 - 2\lambda = 0$$
$$\Rightarrow \lambda(\lambda - 2) = 0$$
$$\lambda = 2 \quad [\because \lambda \neq 0]$$

$$P(X=1 \text{ or } X=0) = P(X=1) + P(X=0)$$
$$= f(1) + f(0)$$
$$= e^{-\lambda}\lambda + e^{-\lambda}$$
$$= e^{-\lambda}(\lambda+1) = 3e^{-2}$$
$$= \frac{3}{e^2}$$

Again, Mean of the distribution
$$= \lambda = 2.$$

Q2 On an average, let 2 cars enter to a parking space for a minute. What is the probability that during any given minute, 4 or more cars will enter to the parking space?

Soln : Let us define a discrete RV.

X : The no. of cars entered into the parking space.

Given, $\lambda = 2$.

$$P(X \geq 4) = 1 - P(X < 4)$$

$$= 1 - \{ P(X=0) + P(X=1) + P(X=2) + P(X=3) \}$$

$$= 1 - \{ e^{-\lambda} + e^{-\lambda}\lambda + \frac{e^{-\lambda}\lambda^2}{2} + \frac{e^{-\lambda}\lambda^3}{6}]$$

$$= 1 - e^{-2} [1 + 2 + 2 + \frac{4}{3}]$$

$$= 1 - \frac{19}{3e^2}$$

H/W

Q1 A car-hire firm has 2 cars which it hires out day-by-day, then the number of demands for each car on each day in which it is distributed as a Poisson distribution. with avg. number of demand per day is 1.5. Calculate the probability ~~on which~~ (proportion of days) on which neither car is used and some demand is refused.

Q2 Find the mean and variance of P.D. by using MGF.

Sol^n 2: Mean using MGF

$$\mu = E(X) = \frac{d}{dt} M_x(t) \Big|_{t=0}$$

$$M_x(t) = E\left(e^{tx}\right) = \sum_{x=0}^{n} e^{tx} f(x)$$

$$= \sum_{x=0}^{n} \frac{e^{tx}\, e^{-\lambda} \lambda^{x}}{x!}$$

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

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

$$= e^{-\lambda} e^{\lambda e^{t}}$$

$$\therefore \mu = \frac{d}{dt} e^{-\lambda} e^{\lambda e^{t}} \Big|_{t=0}$$

$$= \left[e^{-\lambda} \lambda e^{t} e^{\lambda e^{t}} \right]_{t=0}$$

$$= \cancel{e^{-\lambda}} \lambda \cancel{e^{\lambda}}$$

$$= \lambda$$

Variance using MGF

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

$$E(X^2) = \frac{d^2}{dt^2} M_x(t) \Big|_{t=0}$$

$$= \frac{d}{dt} \left(e^{-\lambda} \lambda e^{t} e^{\lambda e^{t}} \right) \Big|_{t=0}$$

$$= \lambda e^{-\lambda} \left[e^{t} e^{\lambda e^{t}} + e^{t} \lambda e^{t} e^{\lambda e^{t}} \right]_{t=0}$$

$$= \lambda e^{-\lambda} \left[e^{\lambda} + \lambda e^{\lambda} \right]$$

$$= \lambda e^{-\lambda} (1+\lambda) e^{\lambda} = \lambda(1+\lambda)$$

$$\therefore \sigma^2 = \lambda(1+\lambda) - \lambda^2 = \lambda.$$

Normal Distribution :

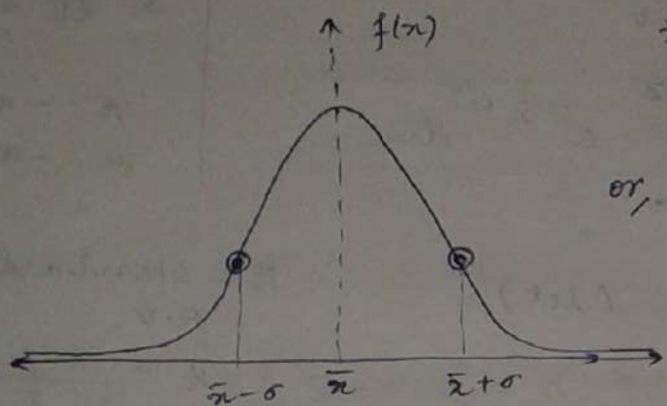
18/10/16

Let X be a continuous type R.V.

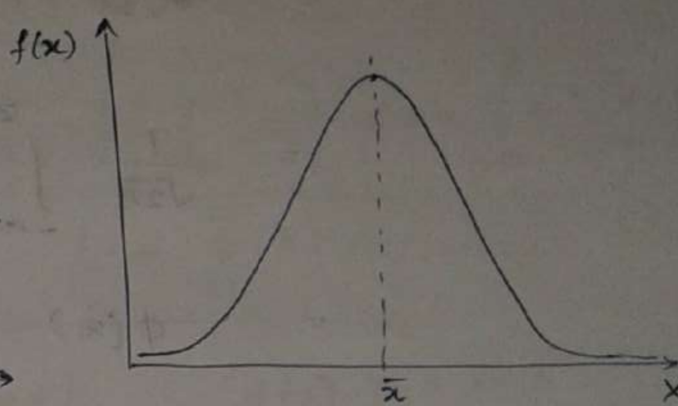
Let us define the prob. density function as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Now X along with above f defined f(x) is called Normal distribution.



or,



Properties

i) It is a continuous type prob. distribution.

ii) Symmetric about $\bar{x}$.

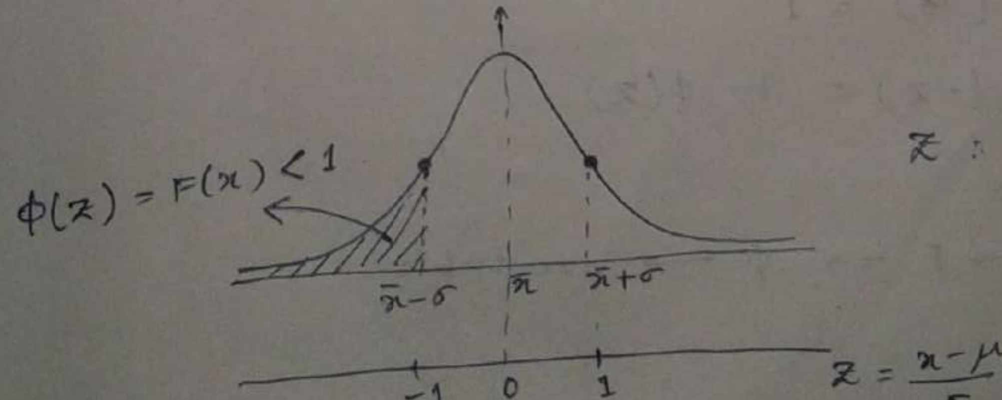
iii) Mean = Median = Mode

iv) $\bar{x}+\sigma$, $\bar{x}-\sigma$: Points of inf. / saddle point.

v) The +ve and -ve x-axis will behave like asymptotes.

vi) The area under the entire curve is 1.

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



Z : standardised R.V.

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

Distribution function

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

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}u^2}\, du$$

Let $\frac{t-\mu}{\sigma} = u$

$\Rightarrow dt = \sigma du$

$t : -\infty \to x$

$u : -\infty \to \frac{x-\mu}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}u^2}\, du$$

$$= \phi(z) \quad \text{(let)}$$

Z : standardized R.V.

$z = \frac{x-\mu}{\sigma}$

Properties

i) $F(x) = \phi(z) = \phi\left(\frac{x-\mu}{\sigma}\right)$

ii) $P(a < X \leq b)$

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

$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

iii) $\phi(0) = \frac{1}{2}$

iv) $\phi(-\infty) = 0$

v) $\phi(\infty) = 1$

vi) $\phi(-z) = 1 - \phi(z)$

$P \to F \to \phi \to$ use table.

Q. Mathematically show that $\phi(-z) = 1 - \phi(z)$

We know, $\phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-\frac{1}{2}u^2}\, du$

$u = -v$

$du = -dv$

$u : -\infty \to -z$

$v : \infty \to z$

$$= -\frac{1}{\sqrt{2\pi}} \int_{\infty}^{z} e^{-\frac{1}{2}v^2}\, dv$$

$$= \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-\frac{1}{2}v^2}\, dv$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} e^{-\frac{1}{2}v^2}\, dv - \int_{-\infty}^{z} e^{-\frac{1}{2}v^2}\, dv \right]$$

$$= 1 - \phi(z).$$

eg. Find $P(\bar{x} - \sigma < X \leq \bar{x} + \sigma)$, $\phi(1) = 0.8413$

$$= F(\bar{x} + \sigma) - F(\bar{x} - \sigma)$$

$$= \phi\left(\frac{\bar{x} + \sigma - \mu}{\sigma}\right) - \phi\left(\frac{\bar{x} - \sigma - \mu}{\sigma}\right)$$

$$= \phi(1) - \phi(-1)$$

$$= \phi(1) - [1 - \phi(1)]$$

$$= 2\phi(1) - 1$$

$$= 0.6826$$

$P(-z < X \leq z)$

$= \phi(z) - \phi(-z)$

$= 2\phi(z) - 1$

$P(\bar{x} - 2\sigma < X \leq \bar{x} + 2\sigma)$

$$= \phi(2) - \phi(-2)$$

$$= 2\phi(2) - 1$$

$$= 1.9544 - 1$$

$$= 0.9544$$

$P(\bar{x} - 3\sigma < X \leq \bar{x} + 3\sigma)$

$$= \phi(3) - \phi(-3)$$

$$= 2\phi(3) - 1$$

$$= 1.9974 - 1$$

$$= 0.9974$$

$.9987 \times 2 = 1.9974$

Theorem : 25/10/16

(Transformation of Mean and Variance)

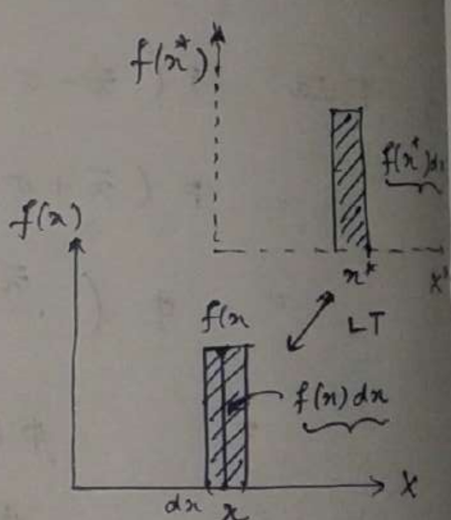
Let X and X^* be two R.V. linearly related by the relationship $X^* = a + bX$

Let X has mean μ and variance σ^2

Then the R.V. X^* has mean μ^* and variance σ^{*2} as follows :

$$\boxed{\mu^* = a + b\mu \quad \Big| \quad \sigma^{*2} = b^2\sigma^2}$$

Moreover, the standardized R.V. $Z = \dfrac{X-\mu}{\sigma}$ has mean 0 and variance 1.



Proof :-

Mean

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

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

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

$f(x^*)\, dx^* = f(x)\, dx$ (area same)

$$= a + b\mu.$$

Variance

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

$$= \int_{-\infty}^{\infty} (\cancel{a} + bx - \cancel{a} - b\mu)^2 f(x)\, dx$$

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

$$= b^2\sigma^2$$

Moreover, given $z = \frac{x-\mu}{\sigma}$

$$z = -\frac{\mu}{\sigma} + \frac{1}{\sigma}x$$

$$\Rightarrow \quad z = a + bx$$

where $a = -\frac{\mu}{\sigma}$, $b = \frac{1}{\sigma}$.

By above theorem, the mean and S.D. of z are given by,

$$\text{Mean} = a + b\mu = -\frac{\mu}{\sigma} + \frac{1}{\sigma}\mu = 0$$

$$\text{S.D.} = b\sigma = \frac{1}{\sigma} \cdot \sigma = 1.$$

or $\quad \text{Variance} = b^2\sigma^2 = \frac{1}{\sigma^2}\sigma^2 = 1.$

<u>Q1</u> Determine the following probabilitys where X is normal.

Given, 0.8 as mean and 4 as variance

$\phi(0.82) = 0.7939$

$\phi(0.1) = 0.5398$

$\phi(0.6) = 0.7257$

i) $P(X \leq 2.44)$

ii) $P(X \leq -1.16)$

iii) $P(X \geq 1)$

iv) $P(2 \leq X \leq 10)$.

Sol^n:

i) $P(X \leq 2.44) = F(2.44) = \phi\left(\frac{2.44 - 0.8}{2}\right)$

$= \phi(0.82)$

$= 0.7939$

ii) $P(X \leq -1.16) = F(-1.16) = \phi\left(\frac{-1.16 - 0.8}{2}\right)$

$= \phi(-0.98)$

$= 1 - \phi(0.98)$

$= 1 - 0.8365$

$=$ ~~0.2745~~ 0.1635

$=$

iii) $P(X \geq 1)$

$= 1 - P(X \leq 1)$

$= 1 - F(1)$

$= 1 - \phi\left(\frac{1-0.8}{2}\right) = 1 - \phi(0.1)$

$= 1 - 0.5398$

$= 0.4602$

iv) $P(2 < X \leq 10)$

$= F(10) - F(2)$

$= \phi(4.6) - \phi(0.6)$

$= 1 - 0.7257$

$= 0.2743.$

$\begin{array}{r} 1.0000 \\ 0.7257 \\ \hline 0.2743 \end{array}$

Q2. Let X is normal with mean 0 and variance 1. Determine the constant 'c' such that

(i) $P(X < c) = 5\%.$

(ii) $P(-c \leq X \leq c) = 99\%.$

Given, $\mu = 0$, $\sigma = 1$.

Sol^n: (i) $P(X \leq c) = 0.05$

$\Rightarrow F(c) = 0.05$

$\Rightarrow \phi\left(\frac{c-0}{1}\right) = 0.05$

$\Rightarrow c = \phi^{-1}(-0.05)$

$= -1.645$

(ii) $P(-c \leq x \leq c) = 99\%$

$\Rightarrow F(c) - F(-c) = 0.99$

$\Rightarrow \phi\left(\frac{c-0}{1}\right) - \phi\left(\frac{-c-0}{1}\right) = 0.99$

$\Rightarrow \phi(c) - \{1 - \phi(c)\} = 0.99$

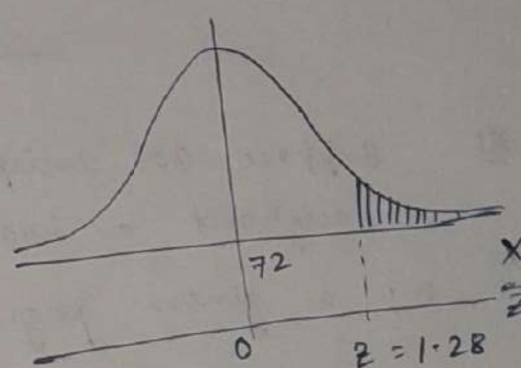
$\Rightarrow 2\phi(c) = 1.99$

$\phi(c) = \frac{1.99}{2}$

$c = \phi^{-1}\{0.995\} = 2.576.$

Q3 ~~The marks x~~ Determine the min. mark a student must get in order to receive an A grade, if the top 10% of students are awarded A grades in an examination where the mean mark is 72 and SD is 9.

Given, $\phi^{-1}(0.9) = 1.28$



$\cancel{\phi^{-1}(0.9) \neq 1}$

$\phi(z) = 0.9$

$\Rightarrow z = \phi^{-1}(0.9)$

$= 1.28$

$z = \frac{x - \mu}{\sigma}$

$\Rightarrow x = z\sigma + \mu$

$= 83.5$

Q Find mean and SD of a normal distribu[tion] in which 7% of items are under 35 a[nd] 89% under 63.

$$\bar{x} - 1.48\sigma = 35$$

$$\bar{x} + 1.23\sigma = 63$$

$$\bar{x} = 50.3$$

$$\sigma = 10.3$$

01/11/1[6]

Q1 Suppose 300 misprints are distributed randomly throughout a book of 500 pages. Find the probabil[ity] P: a given page contains (i) exactly 2 misprints (ii) 2 or more mispri[nts]

Sol^n: Here, $n = 500$

$$\lambda = \frac{300}{500} = 0.6$$

~~$\lambda = np = 500 \times 0.6 = 300$~~

X : No. of misprints in a page

$$x = 2$$

(i) $P(X=2) = f(2) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$$= \frac{e^{-0.6}(0.6)^2}{2!} = 0.09[88]$$

(ii) $P(X \geq 2) = 1 - P(X < 2)$

$$= 1 - \{ P(X=0) + P(X=1) \}$$

$$= 1 - \{ e^{-0.6} + e^{-0.6}(0.6) \}$$

$$= 1 - e^{-0.6}(1.6) = 0.122$$

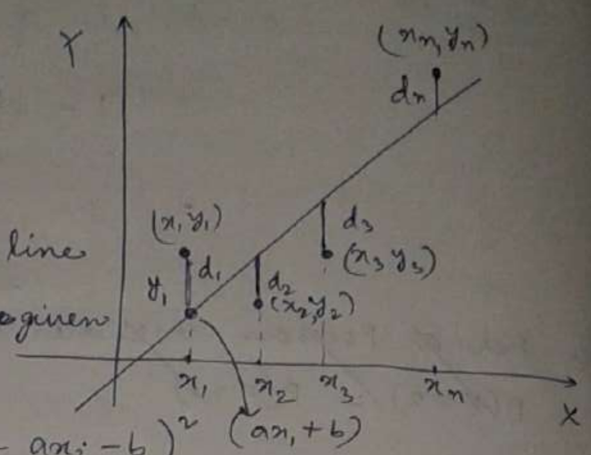
Q4 For a Poisson variate (variate ≡ R.V.) ~~A~~

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

Curve - Fitting :-

01/11/16

Method of Least square Approximation :—



Let $y = ax+b$ be a st. line that best fits to the given n pairs of data.

$$R = \sum_i d_i^2 = \sum_i (y_i - ax_i - b)^2 \longrightarrow (1)$$

As given in the fig, the sum of squares of the distances (R, say) is given by eq^n (1).

Now we need to find the values of a, b to define the straight line subject to the minimization of R i.e. Min R.

To find the minima of R, we have

$\dfrac{\partial R}{\partial a} = 0$	$\dfrac{\partial R}{\partial b} = 0$
$\Rightarrow -2\sum x_i (y_i - ax_i - b) = 0$	$\Rightarrow -2\sum (y_i - ax_i - b) = 0$
$\Rightarrow \sum x_i y_i = a\sum x_i^2 + b\sum x_i \longrightarrow (2)$	$\Rightarrow \sum y_i = a\sum x_i + bn \longrightarrow (3)$

Equations (2) & (3) are called NORMAL equations to the st. line, which need to solve to find the values of a & b, and to be put in the st. line eq^n $y = ax+b$.

(Q) Derive the normal equations for fitting a parabola of the form $y = ax^2 + bx + c$ for a given n pairs of data $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$

Solⁿ :

Q1 By the method of Least square fit a straight line for the following data :

x	1	2	3	4	5
y	14	27	40	55	68

Solⁿ; $n = 5$

Let $y = ax + b$ be a straight line that best fits to the given data.

The normal eqⁿs are :

$$\Sigma y_i = a \Sigma x_i + 5b \quad \text{——— (1)}$$

$$\Sigma x_i y_i = a \Sigma x_i^2 + b \Sigma x_i \quad \text{——— (2)}$$

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
$\Sigma x_i = 15$	$\Sigma y_i = 204$	$\Sigma x_i y_i = 748$	$\Sigma x_i^2 = 55$

$$① \Rightarrow \quad 204 = 15a + 5b$$

$$② \Rightarrow \quad 748 = 55a + 15b$$

$$a = \frac{68}{5}, \quad b = 0$$

$$= 13.6$$

$$\therefore \boxed{y = 13.6x}$$

08/11/16.

(ii) <u>Fitting of a parabola</u> :

Let $y = ax^2 + bx + c$ be the best fit parabola for the given n pairs of data $(x_i\, y_i)$ where $i = 1, 2, 3 \ldots$

Q. Fit a parabola to the following data :

x	0	1	2	3	4
y	1	5	10	22	38

Sol^n:

x_i	y_i	$x_i y_i$	x_i^2	x_i^3	x_i^4
0	1				
1	5				
2	10				
3	22				
4	38				

(iii) Fitting of an Exponential Curve :

Types

a) $y = bx^{a}$, $b > 0$

b) $y = ae^{bx}$, $a > 0$

c) $y = ka^{bx}$, $k, a > 0$

Hint to solve Take logarithm on both sides to get a linear equation which can be treated as $y = ax + b$ form.

eg a) $y = bx^{a}$

$$\log y = \log b + a \log x$$

$$\Rightarrow \quad Y = B + AX \longrightarrow (1)$$

The normal eqn for the st. line (1) is,

$$\sum_{i=1}^{n} Y_i = Bn + A \sum_{i=1}^{n} X_i \longrightarrow (2)$$

& $$\sum_{i=1}^{n} X_i Y_i = B \sum_{i=1}^{n} X_i + A \sum_{i=1}^{n} X_i^2 \longrightarrow (3)$$

x_i	y_i	$X_i = \log x_i$	$Y_i = \log y_i$	X_i^2	X_iY_i

Solve (2) & (3) to get A and B.

$\Rightarrow \quad a = A$

and $b = \text{antilog}(B)$.

Q: Using the method of least square, fit a curve of the form $y = ab^x$, $ab > 0$ to the following data.

x	1	2	3	4
y	4	11	35	100

x_i	y_i	$Y_i = \log y_i$	x_i^2	$x_i Y_i$
1	4		1	
2	11		4	
3	35		9	
4	100		16	
$\Sigma x_i =$		$\Sigma Y_i =$	$\Sigma x_i^2 =$	$\Sigma x_i Y_i =$

$y = ab^x$

$\log y = \log a + x \log b$

$Y = B + Ax$

$\Sigma Y_i = Bn + A \Sigma x_i$

$\Sigma x_i Y_i = B \Sigma x_i + A \Sigma x_i^2$

Correlation and Regression :-

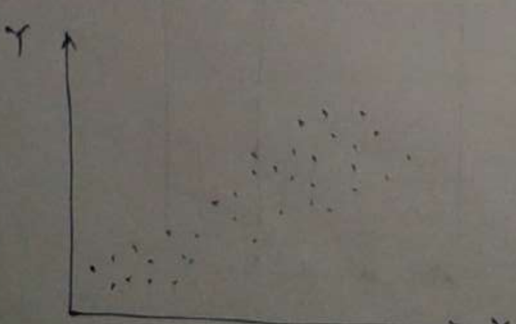
Scatter Diagram : The diagrammatic representation of bivariate data (x, y)

$(X, Y) \rightarrow$ Bivariate ($\downarrow$ 2, $\downarrow$ R.V.)

$\downarrow \quad \downarrow$

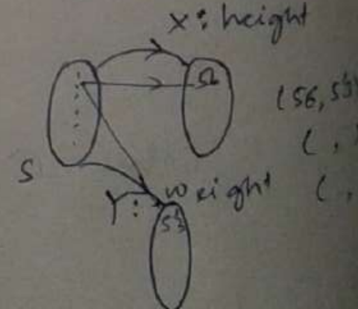
$(x, y) \rightarrow$ B. Data

(i)

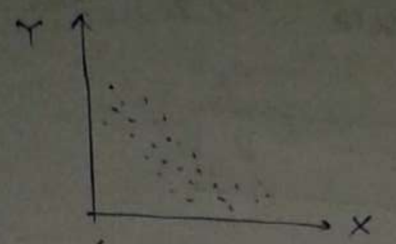


+vely correlated

(X increases ⇒ Y tendency to increase)

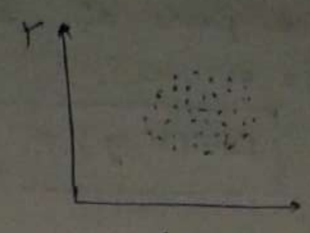


(ii)



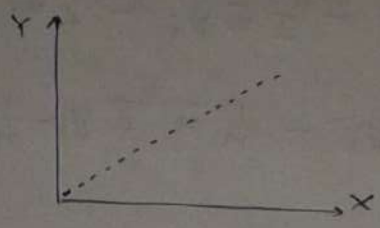
(−vely correlated)
(X increases ⇒ Y has tendency to decrease)

(iii)



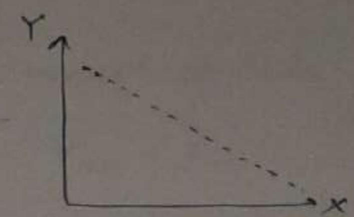
Uncorrelated.

(iv)



Strongly +vely correlated
(X increases ⇒ Y increases)

(v)



Strongly −vely correlated
(X increases ⇒ Y decreases)

<u>Covariance</u> : Let (X, Y) be a B.V with B.V. data (x, y). The covariance of (x, y) for n pairs of data $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ is given by

$$\boxed{Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}$$

<u>Q</u> Find the covariance for the following data:

	x_1	x_2	x_3	x_4	x_5
x	−2	−1	0	1	2
y	4	1	0	1	4

$$Cov(x, y) = \frac{1}{5} \sum_{i=1}^{5} (x_i - 0)(y_i - 2)$$

$$= \frac{1}{5} \left[(-2)(4-2) + (-1)(1-2) + 0 + (1-2) + 2(4-2) \right]$$

$$= \frac{1}{5} \left[-4 + 1 - 1 + 4 \right]$$

$$= 0$$

Thm: For n pairs of bivariate data $(x_i\, y_i)$ for all $i = 1, 2, 3 \ldots n$,

$$\boxed{Cov(x,y) = \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}}$$

We know,

$$Cov(x,y) = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n}\left[\sum x_i y_i - \sum \bar{x} y_i - \sum x_i \bar{y} + \sum \bar{x}\bar{y}\right]$$

$$= \frac{1}{n}\sum x_i y_i - \frac{1}{n}\bar{x}\sum y_i - \frac{1}{n}\bar{y}\sum x_i + \frac{1}{n}\sum \bar{x}\bar{y}$$

$$= \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$$

$$= \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}$$

Corelation: Let (X, Y) be a bivariate with b.v. d[illegible] (x, y). The degree of association or the strength of relationship between x and y is called as the correlation of (x, y).

Mathematically, the correlation of (x, y) can be ~~defined as~~ measured with the help of the correlation coefficient r_{xy}, which is defined as

$$\boxed{r_{xy} = \frac{Cov(x,y)}{\sigma_x\,\sigma_y}}$$

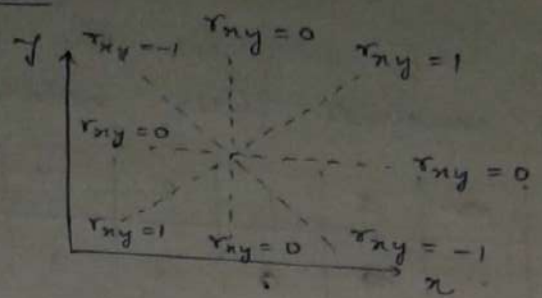
where, $Cov(x,y)$: covariance of x and y

σ_x : S.D. of x

σ_y : S.D. of y

Properties —

(i)



(ii) $r_{xy} = r_{yx}$

(iii) $r_{xy} = 0$ if x, y are uncorrelated.

(v) $r_{x(-x)} = -1$ (iv) $r_{xx} = 1$

Pf (ii) $r_{xy} = \dfrac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$

$$r_{yx} = \frac{\text{Cov}(y, x)}{\sigma_y \sigma_x} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = r_{xy}.$$

$$\text{cov}(y, x) = \frac{1}{n} \sum y_i x_i - \bar{y}\bar{x} = \frac{1}{n} \sum x_i y_i - \bar{x}\bar{y} = \text{cov}(x, y)$$

Pf (iv) $r_{xx} = \dfrac{\text{cov}(x, x)}{\sigma_x \sigma_x} = \dfrac{\frac{1}{n} \sum (x_i - \bar{x})^2}{\sigma_x^2} = 1.$

Pf (v) $r_{x(-x)} = \dfrac{\text{cov}(x, -x)}{\sigma_x \sigma_{-x}} = \dfrac{\frac{1}{n} \sum (x_i - \bar{x})^2}{-(\sigma_x^2)} = -1.$

15/11/16

(8) Find the correlation coefficient for the following bi-variate data and conclude their relationship.

x :	-6	-4	-3	-1		1	2	4	7
y :	-4	-3	-1	-1		0	2	3	4

Sol^n: We know, $r_{xy} = \dfrac{Cov(x,y)}{\sigma_x \sigma_y}$ ------(i)

Now, $Cov(x,y) = \frac{1}{8} \sum_{i=1}^{8} (x_i - 0)(y_i - 0)$

$= \frac{1}{8} \sum_{i=1}^{8} x_i y_i$

$= \frac{1}{8}(84) = \frac{21}{2}$

$\sigma_x^2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2 = \frac{1}{n} \Sigma x_i^2$

$\Rightarrow \sigma_x = \sqrt{\frac{1}{n} \Sigma x_i^2} = \sqrt{\frac{33}{2}}$

$\sigma_y = \sqrt{\frac{1}{n} \Sigma y_i^2} = \sqrt{7}$

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
-6	-4	24	36	16
-4	-3	12	16	9
-3	-1	3	9	1
-1	-1	1	1	1
1	0	0	1	0
2	2	4	4	4
4	3	12	16	9
7	4	28	49	16
$\bar{x} = 0$	$\bar{y} = 0$	$\Sigma x_i y_i = 84$	$\Sigma x_i^2 = \frac{33}{2}$	$\Sigma y_i^2 = 7$

$\therefore$ (i) $\Rightarrow$ $r_{xy} = \dfrac{21/2}{\sqrt{\frac{33}{2}}\ \sqrt{7}}$

$= 0.977 \longrightarrow$ strongly correlated.

CONCLUSION : The bivariate data is strongly correlated since the value of r_{xy} is near to '1'.

__Regression__ : Let (x, y) be a bivariate data of the bivariate (X, Y), the estimation / prediction of ~~one~~ ~~varc~~ the average value of one variable corresponding to a particular value of a number is called Regression.

__eg__

Maths :	35	55	64	39	78	55
Phy :	64	?	64	38	93	?

__Regression line Equation__ :

① R.L 'of y on x'
(used when we need to estimate the value of y for a given x)

② R.L. 'of x on y'
(used ...... x for a given y)

__Thm__ : Let (X, Y) be a bivariate having n pairs of bivariate data $(x_1, y_1)\ (x_2, y_2)\ (x_3, y_3) \ldots (x_n, y_n)$

Let $x_1, x_2, \ldots x_n$ has mean $\bar{x}$ and SD σ_x

and, $y_1, y_2, \ldots y_n$ has mean $\bar{y}$ and SD σ_y.

Then, (i) The Regression Line 'of y on x' is given by

$$(y - \bar{y}) = b_{yx}\,(x - \bar{x})$$

where $b_{yx} = \dfrac{\sigma_y}{\sigma_x}\, r_{xy}$

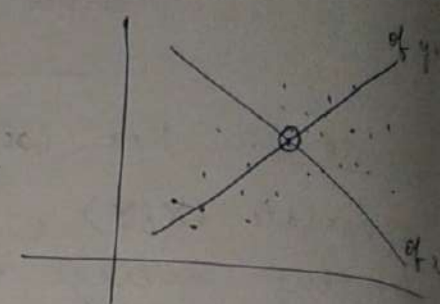
(ii) The R.L. 'of x on y' is given by

$$(x-\bar{x}) = b_{xy}\,(y-\bar{y}),$$

where $b_{xy} = \dfrac{\sigma_x}{\sigma_y}\, r_{xy}$

<u>Proof</u>

Let $y = a+bx$ ⓛ be the best fit regression line of 'y on x' for the given n-pairs of data.



The normal equations are,

$$\left.\begin{aligned} \sum_n y_i &= an + b\sum_n x_i \\ \sum_n x_i y_i &= a\sum_n x_i + b\sum_n x_i^2 \end{aligned}\right\} \longrightarrow ②$$

Dividing eq^n ② by n on both sides,

$$\frac{1}{n}\sum y_i = a + \frac{b}{n}\sum x_i$$

and, $\dfrac{1}{n}\sum x_i y_i = \dfrac{a}{n}\sum x_i + \dfrac{b}{n}\sum x_i^2$

$$\left.\begin{aligned} \therefore\ \bar{y} &= a + \bar{x} b \\ \text{and, } \frac{1}{n}\sum x_i y_i &= a\bar{x} + \frac{b}{n}\sum x_i^2 \end{aligned}\right\} \longrightarrow ③$$

From the above eq^{ns} ③,

$$\frac{1}{n}\sum x_i y_i = (\bar{y} - b\bar{x})\,\bar{x} + \frac{b}{n}\sum x_i^2$$

$$\Rightarrow \frac{1}{n}\sum x_i y_i = \bar{x}\,\bar{y} + b\left[\frac{1}{n}\sum x_i^2 - \bar{x}^2\right]$$

$$\Rightarrow b = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{Cov(x}{\sigma_x^2}$$

$$= \frac{r_{xy}\,\sigma_x\,\sigma_y}{\sigma_x^2}$$

$$\therefore\ b_{yx} = r_{xy}\,\frac{\sigma_y}{\sigma_x}$$

$$\therefore a = \bar{y} - b\bar{x}$$
$$= \bar{y} - \frac{Cov(x,y)}{\sigma_x^2}\bar{x}$$

$$y = a + bx$$
$$= \bar{y} - r_{xy}\frac{\sigma_y}{\sigma_x}\bar{x} + r_{xy}\frac{\sigma_y}{\sigma_x}x$$
$$\Rightarrow y - \bar{y} = (x-\bar{x})\, r_{xy}\frac{\sigma_y}{\sigma_x}$$
$$\therefore \boxed{y - \bar{y} = (x-\bar{x})\, b_{yx}}, \text{ where } b_{yx} = r_{xy}\frac{\sigma_y}{\sigma_x}$$

Similarly, the second part of the theorem can be proved by assuming the regression line of 'x on y' as $x = a + by$

<u>NOTE</u> : b_{yx} is called 'Regression Coefficient' of y on x, and is defined as $b_{yx} = \frac{\sigma_y}{\sigma_x} r_{xy}$

Similarly, b_{xy} is called 'R.C' of 'x on y' and is defined as $b_{xy} = \frac{\sigma_x}{\sigma_y} r_{xy}$

<u>Properties</u> :

1/ r_{xy}, b_{xy}, b_{yx} are of the same sign

2/ Product of the regression coefficients is equal to (Cor. Coeff.)^2

$$b_{xy} \cdot b_{yx} = (r_{xy})^2 \geq 0$$

3/ Both the regression coefficients are of same sign.

4/ Both the regression lines intersect each other at the point $(\bar{x}, \bar{y})$.

5/ The gradient of regression line of 'y on x' $= b_{yx}$
and the " " " " 'x on y' $= \frac{1}{b_{xy}}$

(8) For two variables x and y, the regression lines are

$x + 4y + 3 = 0$ and $4x + 9y + 5 = 0$

(i) Identify, which one is the RL of 'y on x'

(ii) Find the mean of x and y.

(iii) Find the corr. coeff. (r_{xy})

(iv) Estimate the value of x when $y = 1.5$

Sol^n:

$$x + 4y + 3 = 0 \quad \text{——} (1) \qquad 4x + 9y + 5 = 0 \quad \text{——} (2)$$

(i) Let us consider eq^n (1) be the RL of 'y on x'

$$\therefore \quad b_{yx} = -\frac{1}{4}$$

Similarly, $\dfrac{1}{b_{xy}} = -\dfrac{4}{9}$

Now, $b_{yx} \cdot b_{xy} = r_{xy}^2$ $\qquad -1 \leq r_{xy} \leq 1$

$$\therefore \quad r_{xy}^2 = \left(\frac{9}{16}\right) \qquad 0 \leq r_{xy}^2 \leq 1$$

$$r_{xy} = \frac{3}{4} < 1$$

Contradiction:

$$b_{yx} = -\frac{4}{9}$$

$$\frac{1}{b_{xy}} = -\frac{1}{4}$$

$$b_{yx} \cdot b_{xy} = \frac{16}{9} = (r_{xy})^2$$

$$r_{xy} = \frac{4}{3} > 1$$

$\therefore$ eq^n (1) is RL of 'y on x'

and eq^n (2) is RL of 'x on y'.

Two Dimensional Random Variable (≡ BiVariate) 19/11/16

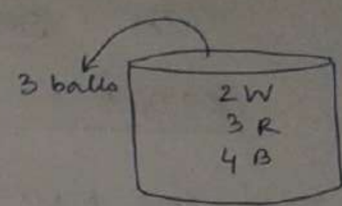
Exercise :

1. A bag contains 2 white, 3 red and 4 black balls. 3 balls are drawn at random without replacement. Let X : be the RV of getting a white ball, and Y : " getting a red ball. Find the probabilities of all combination of (X, Y).

Soln

X \ Y	0	1	2	3
0	1/21	3/14	1/7	1/84
1	1/7	2/7	1/14	0
2	1/21	1/28	0	0

← represents pmf.



$$P(\text{no W, no R}) = P(\text{Black}) = \frac{4}{9} \times \frac{3}{8} \times \frac{2}{7} = \frac{1}{21}$$

$$P(\text{no W, 1R}) = \frac{3}{9} \times \frac{4}{8} \times \frac{3}{7} = \frac{1}{14}$$

$$P(\text{no W, 2R}) = \frac{3}{9} \times \frac{2}{8} \times \frac{4}{7} = \frac{1}{21}$$

$$P(\text{no W, 3R}) = \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} = \frac{1}{84}$$

$$P(\text{1W, 0R}) = \frac{2}{9} \times \frac{4}{8} \times \frac{3}{7} = \frac{1}{21}$$

$$P(\text{2W, 0R}) = \frac{2}{9} \times \frac{1}{8} \times \frac{4}{7} = \frac{1}{63}$$

$$P(\text{1W, 1R}) = \frac{2}{9} \times \frac{3}{8} \times \frac{4}{7} = \frac{1}{21}$$

$$P(\text{1W, 2R}) = \frac{2}{9} \times \frac{3}{8} \times \frac{2}{7} = \frac{1}{42}$$

$$P(\text{2W, 1R}) = \frac{2}{9} \times \frac{1}{8} \times \frac{3}{7} = \frac{1}{84}$$

$$P(X=0, Y=0) = \frac{{}^4C_3}{{}^9C_3}$$

$$P(X=0, Y=1) = \frac{{}^3C_1\ {}^4C_2}{{}^9C_3}$$

$$P(X=0, Y=2) = \frac{{}^3C_2\ {}^4C_1}{{}^9C_3}$$

$$P(X=0, Y=3) = \frac{{}^3C_3}{{}^9C_3}$$

$$P(X=1, Y=0) = \frac{{}^2C_1\ {}^4C_2}{{}^9C_3}$$

$$P(X=1, Y=1) = \frac{{}^2C_1\ {}^3C_1\ {}^4C_1}{{}^9C_3}$$

$$P(X=1, Y=2) = \frac{{}^2C_1\ {}^3C_2}{{}^9C_3}$$

$$P(X=2, Y=0) = \frac{{}^2C_2\ {}^4C_1}{{}^9C_3}$$

$$P(X=2, Y=1) = \frac{{}^2C_2\ {}^3C_1}{{}^9C_3}$$

(3)

(A) Joint Probability Distribution

Discrete

(1) **Joint Prob. Mass Function**

X, Y : discrete R.V.

$$P_{ij} = f_{ij} = P(X = x_i, Y = y_j)$$

is called joint PMF if its satisfies the following properties :-

i) $\sum_i \sum_j P_{ij} = 1$

ii) $P_{ij} \geq 0 \quad \forall\ i, j$

NOTE : The bivariate (X, Y) along with the joint PMF P_{ij} is called joint prob. distribution.

(2) **Commutative distributionf^n**

$$F(x, y) = P(X \leq x, Y \leq y)$$

$$= \sum_{y \leq y_j} \sum_{x \leq x_i} P(X = x_i, Y = y_j)$$

Continuous

(1) **Joint Prob. Density Function**

X, Y : continuous R.V.

$$P\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2},\ y - \frac{dy}{2} \leq Y \leq y + \frac{dy}{2}\right)$$

$$= f(x, y)\, dx\, dy$$

Here, $f(x, y)$ is called the joint prob. density function if it satisfies the following properties :-

i) $f(x, y) \geq 0 \quad \forall\ x, y$

ii) $\int_y \int_x f(x, y)\, dx\, dy = 1$

(2) **Commutative distribution**

$$F(x, y) = P(X \leq x, Y \leq y)$$

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

(3) Properties

a/ $F(-\infty, \infty) = 0 = F(-\infty, -\infty)$
$= F(\infty, -\infty)$

b/ $F(\infty, \infty) = 1$

c/ $P(a < X \leq b,\ Y \leq y)$
$= F(b, y) - F(a, y)$

d/ $P(X \leq x,\ c < Y \leq d)$
$= F(x, d) - F(x, c)$

e/ $P(a < X \leq b,\ c < Y \leq d)$
$= F(b, d) + F(a, c)$
$- F(a, d) - F(b, c)$

(3) Properties

— same —
(a/ to e/)

f/ $\dfrac{\partial^2 F}{\partial x \partial y} = f(x, y)$

(B) Marginal Probability Distribution

Discrete

(1) Marginal prob. mass function

$P(X = x_i) = P(X = x_i, Y = y_1)$
or $P(X = x_i, Y = y_2)$
$\vdots$
or $P(X = x_i, Y = y_n)$

$= \sum_j P(X = x_i, Y = y_j)$

$= P_{i*}$

Now, P_{i*} is called the marginal pmf of X

NOTE : (i) The bivariate (X, Y) along with P_{i*} is called marginal prob. distribution.

(ii) Similarly, bivariate (X, Y) along with P_{*j} is called marginal prob. dis. of y where

$P_{j*} = \sum_i P(X = x_i, Y = y_j)$

Continuous

(1) Marginal prob. density function

$P\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2}, -\infty \leq Y \leq \infty\right)$

$$= \int_{-\infty}^{\infty} \int_{x - \frac{dx}{2}}^{x + \frac{dx}{2}} f(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \left(f(x, y)\, dy\right) [x]_{x - \frac{dx}{2}}^{x + \frac{dx}{2}}$$

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

$$= f_X(x)\, dx$$

where $f_X(x)$ is called marginal density f^n of x

and, $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$

NOTE : $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$

(c) Conditional Probability distribution

discrete

(1) Conditional prob. mass f^n

$$P(X = x_i \mid Y = y_j)$$

$$= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{P_{ij}}{P_{*j}}$$

$$= \frac{\text{Joint pmf}}{\text{marginal pmf}}$$

$$= f(x \mid y)$$

$$P(Y = y_j \mid X = x_i) = \frac{P_{ij}}{P_{i*}}$$

$$= f(y \mid x)$$

$f(x|y) \rightarrow$ conditional prob. of x with given y.

$f(y|x) \longrightarrow$ conditional prob. of y with given x

continuous

(1) Conditional prob. density

$$P\left(x - \frac{dx}{2} \leq X \leq x + \frac{dx}{2} \,\middle|\, y - \frac{dy}{2} \leq Y \ldots\right)$$

$$=$$

---

(8) For the BV prob. distribution of (X, Y) given below, find $P(X \leq 1)$ and $P(Y \leq 3)$

X \ Y	1	2	3	4	5	6
0	0	0	1/32	2/32	2/32	3/32
1	1/16	1/16	1/8	1/8	1/8	1/8
2	1/32	1/32	1/64	1/64	0	2/64

Sol^n $P(X \le 1) = P(X=0) + P(X=1)$

$$= \sum_{j=1}^{6} P(X=0, Y=j) + P(X=1, Y=j)$$

$$= \left\{0+0+\frac{1}{32}+\frac{2}{32}+\frac{2}{32}+\frac{3}{32}\right\} + \left\{\frac{1}{16}+\frac{1}{16}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}\right\}$$

$$= \frac{8}{32} + \frac{10}{16} = \frac{2}{8} + \frac{5}{8} = \frac{7}{8}$$

$P(Y \le 3) = P(Y=1) + P(Y=2) + P(Y=3)$

$$= \sum_{i=0}^{2} P(X=i, Y=1) + P(X=i, Y=2) + P(X=i, Y=3)$$

$$= \left(\frac{1}{16}+\frac{1}{32}\right) + \left(\frac{1}{16}+\frac{1}{32}\right) + \left(\frac{1}{32}+\frac{1}{8}+\frac{1}{64}\right)$$

$$= \frac{23}{64}$$

---

(8) In the above question, find the following probabilities :-

(i) $P(X \le 1, Y \le 3)$

(ii) $P(X \le 1 \mid Y \le 3)$

(iii) $P(Y \le 3 \mid X \le 1)$

(iv) $P(X + Y \le 4)$

Sol^n (i) $P(X \le 1, Y \le 3)$

$$= \sum_{y \le 3} \sum_{x \le 1} P(X=x, Y=3)$$

$$= 0 + 0 + \frac{1}{32} + \frac{1}{16} + \frac{1}{16} + \frac{1}{8}$$

$$= \frac{9}{32}$$

(ii) $P(X \le 1 \mid Y \le 3) = \dfrac{P(X \le 1, Y \le 3)}{P(Y \le 3)}$

$$= \frac{9/32}{23/64} = \frac{18}{23}$$

(iii) $P(Y \leq 3 \mid X \leq 1) = \dfrac{P(Y \leq 3, X \leq 1)}{P(X \leq 1)}$

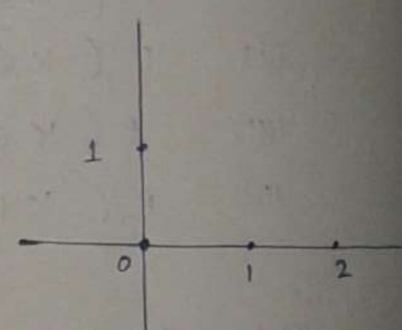
$$= \frac{9/32}{7/8} = \frac{9}{\cancel{32}_{4}} \cdot \frac{\cancel{8}}{7} = \frac{9}{28}$$

---

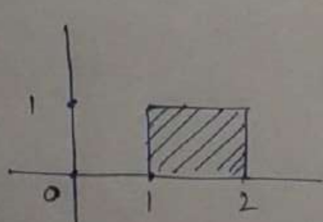
(8) The joint pdf of the bivariate (X, Y) is given by

$$f(x,y) = xy^2 + \frac{x^2}{8}, \quad \begin{array}{l} 0 \leq x \leq 2 \\ 0 \leq y \leq 1 \end{array}$$

Find
(i) $P(X > 1)$
(ii) $P(Y < 1)$
(iii) $P(Y < \frac{1}{2} \mid X > 1)$
(iv) $P(X < Y)$
(v) $P(Y > \frac{1}{2})$
(vi) $P(X + Y < 1)$



(i) $P(X > 1)$



$$\int_0^1 \int_1^2 xy^2 + \frac{x^2}{8} \, dx \, dy.$$

$$= \int_0^1 \left[\frac{x^2 y^2}{2} + \frac{x^3}{24} \right]_1^2 dy$$

$$= \int_0^1 \left(2y^2 + \frac{1}{3}\right) - \left(\frac{y^2}{2} + \frac{1}{24}\right) dy$$

$$= \int_0^1 \frac{3}{2} y^2 - \frac{7}{24} \, dy$$

$$= \frac{y^3}{2} - \frac{7}{24} y \Big|_0^1$$

$$= \frac{19}{24}$$

(ii) $P(Y<1) = \int_0^1 \int_0^2 f(x,y)\,dx\,dy$ = entire volume = 1.

(iii) $P(Y<\frac{1}{2} \mid X>1) = \dfrac{P(Y<\frac{1}{2}, X>1)}{P(X>1)}$

$$= \frac{\frac{5}{24}}{\frac{19}{24}} = \frac{5}{19}.$$

~~(iv) Repeated~~

(v) $P(Y>\frac{1}{2}) = 1 - P(Y<\frac{1}{2})$

$$= 1 - \int_0^{1/2} \int_0^2 xy^2 + \frac{x^2}{8}\,dx\,dy$$

$$= 1 - \int_0^{1/2} \left[\frac{x^2y^2}{2} + \frac{x^3}{24} \right]_0^2 dy$$

$$= 1 - \int_0^{1/2} \frac{3}{2}y^2 - \frac{7}{24}\,dy$$

$$= 1 - \frac{y^3}{2} - \frac{7}{24}y \Big|_0^{1/2}$$

$$= 1 - \left[\frac{1}{16} - \frac{7}{48} \right] = 1 - \left\{ \frac{3}{48} \right\}$$

(iv) $P(X<Y) = \int_0^1 \int_0^{1-y} f(x,y)\,dx\,dy$

(vi) $P(X+Y<1) = \int_0^2 \int_0^y f(x,y)\,dx\,dy$