Lecture - 2

Binomial distribution : Suppose there are n independent trials of an experiment. Let p denote the probability of success and q be the probability of a failure in a single trial. (Here let a random experiment be performed repeatedly, each repetition being called a trial and let the happening of an event in a trial is called a success and its non-happening a failure)

Now the probability of getting r successes in these n independent trials $= (\underbrace{p \cdot p \cdots p}_{r \text{ times}})(\underbrace{q \cdot q \cdots q}_{(n-r) \text{ times}})$.

$= p^r q^{(n-r)}$ {by the multiplication theorem}.

But these r ~~r~~ successes in n trials can occur in ${}^nC_r$ ways. ~~Hence~~ since ${}^nC_r$ ways are mutually exclusive, then the probability of having r ~~r~~ successes out of n trials is given by.

$$P(r) = {}^nC_r\, p^r q^{n-r} \quad \text{(addition th.)}$$

Note : 1) Above probability fun^n is probability mass function (pmf)

$\because P(X=r) \geq 0 \;\; \forall \; r = 0,1,2,\ldots n$ as $p, q \geq 0$

and $\sum_{r=0}^{n} P(X=r) = \sum_{r=0}^{n} {}^nC_r\, p^r q^{n-r}$

$= {}^nC_0\, p^0 q^{n-0} + {}^nC_1\, p q^{n-1} + \cdots + {}^nC_{n-1}\, p^{n-1} q + {}^nC_n\, p^n$

$= (q+p)^n$.

$= 1^n = 1$

2) Since the probabilities $P(X=r)$ i.e. ${}^nC_r\, p^r q^{n-r}$ are the successive terms in the expansion of Binomial expression $(q+p)^n$, therefore this distribution is called Binomial distribution.

3) Binomial distribution is also known as Bernoulli distribution ~~of~~.

Q) If 10% of the pens manufactured by a company are defective. Find the probability that a box of 12 pens contains: (i) exactly two defective pens (ii) at least two defective pens.

Soln. Let p = probability of a defective pen = 0.1

$\therefore q = 0.9$ and $n = 12$.

(i) Probability that the box contains two defective pens.

$= {}^{12}C_2 p^2 q^{10} = {}^{12}C_2 (0.1)^2 (0.9)^{10} = 0.2301$.

(ii) Probability that the box contains at least two defective pens = 1 − [Prob. that the box contains either none or one non-defective pen]

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

$= 1 - [{}^{12}C_0 p^0 q^{12} + {}^{12}C_1 p^1 q^{11}]$

$= 1 - [{}^{12}C_0 (0.9)^{12} + {}^{12}C_1 (0.1)(0.9)^{11}]$

$= 0.341$.

Q.4 The incidence of an occupational disease in an industry is such that the workmen have a 20% chance of suffering from it. What is the probability that out of 6 workmen 4 or more will suffer from disease?

Soln. p = the probability of a man suffering from disease

$= \frac{20}{100}$

$= \frac{1}{5}$

then $q = 1 - p = 1 - \frac{1}{5}$

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

Here $n = 6$.

Required probability that out of 6 workmen 4 or more will suffer from disease $= P(4) + P(5) + P(6)$

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

$= 0.53$
3/25

Note 1. In the binomial distribution n, p are said to be parameters.
2. If $p = q = \frac{1}{2}$, the binomial distribution is called symmetrical distribution otherwise it is called skew distribution.

Mean and Variance of Binomial Distribution :

For the binomial distribution

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

$$\text{Mean} = \mu = E(X) = \sum_{x=0}^{n} x \, P(x)$$

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

$$= 0 + n C_1 q^{n-1} p + 2 \cdot {}^nC_2 \, q^{n-2} p^2 + \dots + n \cdot {}^nC_n \, q^0 p^n$$

$$= n q^{n-1} p + n(n-1) q^{n-2} p^2 + \frac{n(n-1)(n-2)}{2!} q^{n-3} p^3 + \dots + n p^n$$

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

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

$$= np.$$

Variance

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

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

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

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

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

$$\therefore \quad \sigma^2 = \sum_{x=0}^{n} x^2 P(x) - \mu^2$$

$$= \sum_{x=0}^{n} [x + x(x-1)] P(x) - (np)^2$$

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

Scanned by CamScanner

Variance $\sigma^2 = \mu + \sum_{r=0}^{n} r(r-1)\, P(r) - n^2p^2.$

$= \mu + \sum_{r=0}^{n} r(r-1)\, {}^nC_r\, q^{n-r} p^r - n^2p^2$

$= \mu + [2.1\, {}^nC_2\, q^{n-2} p^2 + 3.2\, {}^nC_3\, q^{n-3} p^3 + \dots + n(n-1)\, {}^nC_n\, p^n]$

$- n^2p^2$

$= \mu + [n(n-1) q^{n-2} p^2 + n(n-1)(n-2)\, q^{n-3} p^3$

$+ \dots + n(n-1) p^n] - n^2p^2$

$= \mu + n(n-1)p^2 [q^{n-2} + (n-2) q^{n-3} p + \dots + p^{n-2}] - n^2p^2$

$= \mu + n(n-1)p^2 [\,{}^{n-2}C_0\, q^{n-2} + {}^{n-2}C_1\, q^{n-3} p + \dots$

$+ {}^{n-2}C_{n-2}\, p^{n-2}] - n^2p^2.$

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

$= np - np^2$

$= np(1-p).$

$= npq.$

Q If on an average 8 ships out of 10 arrive safely at a port. Find the mean and standard deviation of the number of ships arriving safely out of a ~~toto~~ total of 1600 ships.

Soln: Here p = probability of safe arrival $= \frac{8}{10} = 0.8.$

$q = 1 - p = 0.2$

Mean of ships returning safely is given by.

$np = 1600 \times 0.8 = 1280.$

standard deviation $= \sqrt{npq} = 16.$

Recurrence Formula for the Binomial distribution:

$\because$ $P(r) = {}^nC_r\, p^r q^{n-r}$

Similarly, $P(r+1) = {}^nC_{r+1}\, p^{r+1} q^{n-r-1}$.

$$\therefore \quad \frac{P(r+1)}{P(r)} = \frac{n!\, p^{r+1} q^{n-r-1}}{(r+1)!(n-r-1)!} \Big/ \frac{n!\, p^r q^{n-r}}{r!\,(n-r)!}$$

$$= \frac{n-r}{r+1}\frac{p}{q}$$

$$\Rightarrow \quad P(r+1) = \frac{(n-r)}{(r+1)} \cdot \frac{p}{q} P(r)$$

Fitting a Binomial distribution: When a binomial distribution is to be fitted to observe data, the following procedure is adopted.

1) Find the values of p and q.
2) Expand the binomial $(q+p)^n$
3) Multiply each term of the expanded binomial by N (the total frequency of the given set of data) in order to obtain the expected frequency in each category.

Q. Six dice are thrown 729 times. How many times do you expect atleast three dice to show a five or six?

Soln: p = the probability of getting 5 or 6 with one dice

$= \frac{2}{6} = \frac{1}{3}$

$q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

$n = 6,\ N = 729$

The expected no. of times atleast three dice showing five or six $= 729\,[P(3) + P(4) + P(5) + P(6)]$

$$= 729\left[{}^6C_3\left(\tfrac{2}{3}\right)^3\left(\tfrac{1}{3}\right)^3 + {}^6C_4\left(\tfrac{2}{3}\right)^2\left(\tfrac{1}{3}\right)^4 + {}^6C_5\left(\tfrac{2}{3}\right)\left(\tfrac{1}{3}\right)^5 + {}^6C_6\left(\tfrac{1}{3}\right)^6\right]$$

$$= 729\left[\frac{160}{729} + \frac{60}{729} + \frac{12}{729} + \frac{1}{729}\right] = 233$$

Q. Out of 800 families with 4 children each, how many families would be expected to have 2 boys and 2 girls.

Soln: Since probabilities for boys and girls are equal.

p = probability of having a boy $= \frac{1}{2}$

q = probability of having a girl $= \frac{1}{2}$.

$n = 4$, $N = 800$

The expected number of families having 2 boys and 2 girls $= 800 \times {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$

$= 800 \times 6 \times \frac{1}{16} = 300.$

Q The following data show the number of seeds germinating out of 10 on damp filter for 80 set of seeds. Fit a binomial distribution to this data

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

Soln:

X ;	0	1	2	3	4	5	6	7	8	9	10	
f ;	6	20	28	12	8	6	0	0	0	0	0	N = 80
fX ;	0	20	56	36	32	30	0	0	0	0	0	$\Sigma fX = 174$

$\bar{X} = \frac{\Sigma fX}{N} = \frac{174}{80}$

$\bar{X} = 2.175$

Mean $= np = 2.175.$

$p = \frac{2.175}{n} = \frac{2.175}{10} = .2175$

$q = 1 - p = .7825$

The theoretical frequencies are given below

X	Theoretical frequencies $N \times {}^nC_r p^r q^{n-r}$
0	$80 \times (.7825)^{10}(.2175)^0 = 6.9$
1	$80 \times {}^{10}C_1(.7825)^9(.2175) = 19.1$
2	$80 \times {}^{10}C_2(.7825)^8(.2175)^2 = 24.0$
3	$80 \times {}^{10}C_3(.7825)^7(.2175)^3 = 17.8$
4	8.6
5	2.9
6	0.7
7	0.1
8	0.0
9	0.0
10	0.0
Total	80.1

Q. Fit a binomial distribution to the following data

x:	0	1	2	3	4	5
f:	2	14	20	34	22	8

Soln Given $n=5$, $N = \Sigma f = 100$

mean of the given frequency distribution

$$= \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{14+40+102+88+40}{100} = 2.84$$

$\because$ mean of binomial distribution $= np$

$\Rightarrow np = 2.84$

$$p = \frac{2.84}{n} = \frac{2.84}{5} = 0.568$$

$q = 1-p = 1-0.568 = 0.432$

The Theoretical frequencies are given as

X	Theoretical frequency. ($N\, {}^nC_x p^x q^{n-x}$)
0	$100 \times (0.432)^5 = 1.5$
1	$100 \times 5 (0.568)(0.432)^4 = 9.8$
2	$100 \times 10 (0.568)^2 (0.432)^3 = 26.0$
3	$100 \times 10 (0.568)^3 (0.432)^2 = 34.2$
4	$100 \times 5 (0.568)^4 (0.432) = 22.5$
5	$100 \times (0.568)^5 = 5.9$
	Total 99.9

Note:

~~Arith Mean~~

(1) Arithmetic mean (AM) : It is a quotient obtain by dividing the sum of observations. (Σx_i) by the number of observations (n) $\left[\bar{x} = \frac{\Sigma x_i}{n}\right]$ i.e. Arithmetic Average.

(2) Median : The median of a series of data is defined as that value which divides the whole series in to two equal parts.

Determination of median : arrange in ascending or descending order.
(for ungrouped data)

If no. of observations n is odd then median is $\left(\frac{n+1}{2}\right)$th observation

If n is even then median is average of $\left(\frac{n}{2}\right)$th and ~~$\frac{n+1}{2}$~~ $\left(\frac{n}{2}+1\right)$th observation.

(for grouped data) : Calculate cumulative frequencies.

If total frequency n is odd then $\left(\frac{n+1}{2}\right)$th observation is median

If n is even then mean of $\frac{n}{2}$th and $\left(\frac{n}{2}+1\right)$th observation is median

(3) Mode : The mode is the value of a variate that occurs most often. i.e. the point having maximum frequency.

Scanned by CamScanner

Binomial Distribution :

The Moment Generating Function about origin is

$$M_X(t) = E(e^{Xt}) = \sum_{r=0}^{n} e^{tx_r} p(r)$$

$$= \sum_{r=0}^{n} e^{tr}\, {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} {}^nC_r\, (pe^t)^r q^{n-r}$$

$$\boxed{M_X(t) = (q+pe^t)^n}$$

Moments about origin

$$\mu_1' = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[n pe^t (q+pe^t)^{n-1}\right]_{t=0} = np$$

$$\boxed{\mu_1' = np}$$

$$\mu_2' = \left[\frac{d^2 M_X(t)}{dt^2}\right]_{t=0} = np\left[e^t(q+pe^t)^{n-1} + (n-1)e^{2t}p\,(q+pe^t)^{n-2}\right]_{t=0}$$

$$= \cancel{np[q^{n-1} + (n-1)pq^{n-2}]} = np[1+(n-1)p] = np + n(n-1)p^2$$

$$\boxed{\mu_2' = n^2p^2 + npq = np(np+q)}$$

$$\mu_3' = \left[\frac{d^3 M_X(t)}{dt^3}\right]_{t=0} =$$

$$\boxed{\mu_3' = np + n(n-1)(n-2)p^3 + 3n(n-1)p^2}$$

$$\mu_4' = \left[\frac{d^4 M_X(t)}{dt^4}\right]_{t=0}$$

$$\boxed{\mu_4' = \cancel{np+} n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np}$$

Central Moments

$$\mu_1 = \sum_{i=1}^{n} p_i (x_i - \bar{x})$$

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

$$\because \mu_2 = \mu_2' - \mu_1'^2 = n^2p^2 + npq - n^2p^2$$

$$\boxed{\mu_2 = npq}$$

$$\because \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\boxed{\mu_3 = npq(q-p)}$$

Scanned by CamScanner

$\because \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

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

Moment Generating function about $\bar{x}$ $(=\mu = np)$

$M_X(t) = E[e^{t(X-\mu)}] = E[e^{(tX-\mu t)}] = e^{-\mu t}E[e^{tX}]$

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

$$\boxed{M_X(t) = (qe^{-pt}+pe^{qt})^n}$$

Moments about mean can be calculated by MGF about $\bar{x}$

$\mu_1 = \left[\frac{d}{dt}M_X(t) \text{ about Mean}\right]_{t=0}$

$\mu_1 = 0$

$\mu_2 = npq$ and so on.

Recurrence Relation for the Central Moments of Binomial Distribution

By definition

$\mu_r = E[(X-\mu)^r] = \sum_{x=0}^{n}(x-np)^r P(x)$

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

$\mu_r = \sum_{x=0}^{n} {}^nC_x (x-np)^r p^x (1-p)^{n-x}$

Now differentiating both sides w.r.to p, we get,

$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^nC_x\Big[(-n)\, r(x-np)^{r-1} p^x(1-p)^{n-x}$
$+ (x-np)^r\{xp^{x-1}(1-p)^{n-x} - p^x(n-x)(1-p)^{n-x-1}\}\Big]$

$= \sum_{x=0}^{n}(-nr)\cdot(x-np)^{r-1}\cdot {}^nC_x\, p^x(1-p)^{n-x}$
$+ \sum_{x=0}^{n}(x-np)^r\, {}^nC_x\, p^x q^{n-x}\left\{\frac{x}{p} - \frac{(n-x)}{q}\right\}$

$= -nr\sum_{x=0}^{n}(x-np)^{r-1}P(x) + \sum_{x=0}^{n}(x-np)^r P(x)\left[\frac{x-np}{pq}\right]$

$\frac{d\mu_r}{dp} = (-nr)\cdot\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$

or $\boxed{\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]}$

Karl Pearson's β and γ coefficients for Binomial Distribution :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[npq(q-p)]^2}{(npq)^3} = \frac{(q-p)^2}{npq}$$

$$\boxed{\beta_1 = \frac{(1-2p)^2}{npq}}$$

$$\boxed{\gamma_1 = \sqrt{\beta_1} = \frac{1-2p}{\sqrt{npq}}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq[1+3(n-2)pq]}{(npq)^2}$$

$$\boxed{\beta_2 = 3 + \frac{1-6pq}{npq}}$$

$$\boxed{\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}}$$

Distribution function of Binomial Distribution :

$$F(x) = P(X \leq x)$$
$$= P(X \leq r) = \sum_{r=-\infty}^{r} {}^nC_r\, p^r q^{n-r}$$

Probability Generating function of Binomial Distribution :

$$G_X(z) = \sum_{i=0}^{n} P_i z^i$$

at $i = r$

$$= \sum_{r=0}^{n} {}^nC_r\, p^r q^{n-r} z^r$$
$$= \sum_{r=0}^{n} {}^nC_r\, (pz)^r q^{n-r}$$

$$\boxed{G_X(z) = (q+pz)^n}$$

Mode of Binomial Distribution : Mode is the value of x for which $P(x)$ is maximum. Let $X = r$ be the modal value.

i.e. $P(X=r) > P(X=r-1)$ and $P(X=r) > P(X=r+1)$

Now $$\frac{P(X=r)}{P(X=r-1)} = \frac{{}^nC_r\, p^r q^{n-r}}{{}^nC_{r-1}\, p^{r-1} q^{n-r+1}} = \frac{n-r+1}{r}\frac{p}{q} > 1$$

$\Rightarrow \quad r < (n+1)p$ —— (1)

Now $\dfrac{P(X=r)}{P(X=r+1)} = \dfrac{{}^nC_r\, p^r q^{n-r}}{{}^nC_{r+1}\, p^{r+1} q^{n-(r+1)}} = \dfrac{r+1}{(n-r)} \dfrac{q}{p} > 1$

$\Rightarrow \quad r > np - q$ —— ②

From ① and ②, $\boxed{np-q < r < (n+1)p}$

or $\{(n+1)p-1\} < r < \{(n+1)p\}$

Tutorial

T.
R.V.

Q1 The probability distribution of a random variable X is given below. Find (i) $E(X)$, (ii) $Var.(X)$, (iii) $E(2X-3)$ (iv) $Var.(2X-3)$

Sol

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

Solⁿ (i) $E(X) = \Sigma x_i p_i = 0 = \bar{x}$

$Var.(X) = E\{(X-\bar{x})^2\} = E(X^2) - \{E(X)\}^2$

$= \Sigma x_i^2 p_i - (\bar{x})^2$

(ii) $Var.(X) = 4\times0.2 + 0.1 + 0 + 0.3 + 0.4 = 1.6$

(iii) $E(2X-3) = 2E(X) - 3$

$= 2\times0 - 3 = -3$

(iv) $Var.(2X-3) = 2^2 Var.(X) = 4(1.6) = 6.4$

Q2. Calculate the first four moments about mean from the following data:

x :	0	1	2	3	4	5	6	7	8
f :	5	10	15	20	25	20	15	10	5

Also, calculate the values of β_1 and β_2.

Solⁿ.

x	f	fx	$(x-\bar{x})$	$f(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
0	5	0	-4	-20	80	-320	1280
1	10	10	-3	-30	90	-270	810
2	15	30	-2	-30	60	-120	240
3	20	60	-1	-20	20	-20	20
4	25	100	0	0	0	0	0
5	20	100	1	20	20	20	20
6	15	90	2	30	60	120	240
7	10	70	3	30	90	270	810
8	5	40	4	20	80	320	1280
	$\Sigma f = 125$	$\Sigma fx = 500$		$\Sigma f(x-\bar{x}) = 0$	$\Sigma = 500$	$\Sigma = 0$	$\Sigma = 4700$

where $\bar{x} = \frac{\Sigma f(x)}{\Sigma f} = \frac{500}{125} = 4$

Scanned by CamScanner

Moments about mean

$\mu_1 = \frac{\Sigma f(x-\bar{x})}{\Sigma f} = 0$; $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$

$\mu_2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} = 4$; $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{37.6}{16} = 2.35$

$\mu_3 = \frac{\Sigma f(x-\bar{x})^3}{\Sigma f} = 0$

$\mu_4 = \frac{\Sigma f(x-\bar{x})^4}{\Sigma f} = 37.6$

Q.3. A continuous random variable X is distributed over the interval [0,1] with pdf $f(x) = ax^2+bx$, where a, b are constants. If the mean of X is 0.5, find the values of a and b.

Soln:

$\because$ f(x) is pdf

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

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

or $\frac{a}{3}+\frac{b}{2} = 1$ or $2a+3b = 6$ —— (1)

$\because$ Given mean $= \bar{x} = 0.5$

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

$0.5 = \int_0^1 x(ax^2+bx)\,dx$

or $0.5 = \frac{a}{4}+\frac{b}{3}$ or $3a+4b = 6$ —— (2)

From (1) and (2), we get

$a = -6$

$b = 6$

Mathematical Expectation :

(a) For univariate variable

∴ The expaction of a random variable X is defined as

$$\bar{X} = E(X) = \begin{cases} \sum_i x_i p_i & ; \text{if } X \text{ is discrete R.V. with pmf } p_i \\ \int_{-\infty}^{\infty} x \cdot f(x)\, dx & ; \text{if } X \text{ is continuous RV with pdf } f(x) \end{cases}$$

provided the relevant sum or integral is absolutely convergent and $\bar{X}$ denotes mean of the corresponding distribution.

If X is a R.V and $g(X)$ is any function of X, then

$$E[g(X)] = \begin{cases} \sum_i g(x_i) p_i & ; \text{if } X \text{ is discrete R.V with } P(X=x_i)=p_i \\ \int_{-\infty}^{\infty} g(x) f(x) dx & ; \text{if } X \text{ is continuous R.V} \end{cases}$$

(b) For Bivariate R.V :

If (X, Y) is a two-dimensional random variable then

$$E\{h(x,y)\} = \sum_i \sum_j h(x_i, y_j)\, p_{ij} \quad ; \text{ for discrete R.V.}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx\, dy \; ; \quad \text{for conti. R.V.}$$

where $f(x,y)$ denotes joint p.d.f.
and p_{ij} denotes joint p.m.f.

Addition Theorem of Expectation : If $X_1, X_2 \ldots, X_n$ are R.V. then

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

Multiplication Theorem of Expectation If $X_1, X_2 \ldots, X_n$ are n independent R.V. then

$$E(X_1 X_2 \ldots X_n) = E(X_1)\, E(X_2) \ldots E(X_n)$$

Note : Th : $E(aX+b) = aE(X) + b$; a, b are constants

Q Two unbiased dice are thrown. Find the expected value of the sum of numbers of pts. on them.

Soln:

~~Values~~ The probability distribution of X (the sum of the numbers obtained on two dice) is

Values of X, x :	2	3	4	5	6	7	8	9	10	11	12
$P(x)$:	$\frac{1}{36}$	$2/36$	$3/36$	$4/36$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$E(X) = \left(2\times\frac{1}{36}\right) + \left(3\times\frac{2}{36}\right) + \left(4\times\frac{3}{36}\right) + \left(5\times\frac{4}{36}\right) + \left(6\times\frac{5}{36}\right)$$
$$+\left(7\times\frac{6}{36}\right) + \left(8\times\frac{8}{36}\right) + \left(9\times\frac{4}{36}\right) + \left(10\times\frac{3}{36}\right)$$
$$+\left(11\times\frac{2}{86}\right) + \left(12\times\frac{1}{36}\right)$$

$$= \frac{252}{36} = 7.$$

Q The probability that there is atleast one error in an accounts statement prepared by A is 0.2 and for B and C they are 0.25 and 0.4 respectively. A, B and C prepared 10, 16 and 20 statements respectively. Find the expected number of correct statements in all.

Soln: Given that $P(A) = .2$, $P(B) = .25$ and $P(C) = .4$
where events A, B, C denotes for an error in accounts prepared by them.

$P(\bar{A}) = 1 - .2 = .8$

$P(\bar{B}) = 1 - .25 = .75$

$P(\bar{C}) = 1 - .4 = .6$

Let X be the random variable which denote number of account statements prepared by them.

Values of X, x :	10	16	20
$p(x)$:	.8	.75	.6

$E(X) = (10\times.8) + (16\times.75) + (20\times.6)$
$= 32$

~~is also known as binomial distribution~~

Q Find the expectation of the number on a dice when thrown

Ans: Let X be the random variable which represents the number on dice when thrown, then its probability distribution is

X :	1	2	3	4	5	6
$P(X)$:	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$$E(X) = \left(1\times\frac{1}{6}\right)+\left(2\times\frac{1}{6}\right)+\left(3\times\frac{1}{6}\right)+\left(4\times\frac{1}{6}\right)+\left(5\times\frac{1}{6}\right)+\left(6\times\frac{1}{6}\right)$$

$$= \frac{7}{2}$$

Note: Th: 1 $E(X_1+X_2+\dots+X_n) = E(X_1)+E(X_2)+\dots+E(X_n)$.

$X_1, X_2, \dots X_n$ are random variables, provided all expectations exists

Th: 2 $E(X_1 X_2 \dots X_n) = E(X_1)\,E(X_2)\dots E(X_n)$.

~~provided all~~ $X_1, X_2, X_3, \dots, X_n$ are Independent random variables and provided all expectations exist.

Add some examples of Expectations for conti. R.V.

Bivariate R.V. (discr, conti)

Variance.

$$Var(X) = \sigma^2 = E[(X_i - \bar{X})^2] = \begin{cases} \sum_i (x_i-\bar{X})^2 p_i & ; X \text{ is discrete R.V} \\ \int_{-\infty}^{\infty} (x-\bar{X})^2 f(x)dx & ; X \text{ is conti. R.V.} \end{cases}$$

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

Note: (i) $Var(aX) = a^2 Var(X)$.
(ii) $Var\cdot(aX+b) = a^2 Var\cdot(X)$.

Q A random variable X has the following probability distribution:

x_i :	-2	-1	0	1	2	3
p_i :	0.1	k	0.2	$2k$	0.3	k

(i) Calculate the mean of X
(ii) Variance of X

Sol^n: $\because \sum p_i = 1$

$\therefore 0.6 + 4k = 1$

$k = 0.1$

x_i :	-2	-1	0	1	2	3	
p_i :	0.1	0.1	0.2	0.2	0.3	0.1	
p_ix_i :	-0.2	-0.1	0	0.2	0.6	0.3	$\sum p_ix_i = 0.8$
$p_ix_i^2$:	0.4	0.1	0	0.2	1.2	0.9	$\sum p_ix_i^2 = 2.8$

$\because E(X) = \sum p_ix_i = 0.8$

$E(X^2) = \sum p_i x_i^2 = 2.8$

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

$\sigma^2 = (2.8) - (0.8)^2$

$\sigma^2 = 2.16$

Q A random variable X have a following p.d.f.

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

Find (i) $E(X)$ (ii) Variance of X
(iii) S.D. of X (iv) $E(3X^2 - 2X)$

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

$$= \int_0^2 x\left(\frac{1}{2}x\right)dx = \left(\frac{x^3}{6}\right)_0^2 = \frac{8}{6} = \frac{4}{3} = \bar{X}$$

(ii) $\sigma^2 = E\left[(x-\bar{X})^2\right] = \int_{-\infty}^{\infty}\left(x-\frac{4}{3}\right)^2 f(x)\,dx$

$$= \int_0^2 \left(x-\frac{4}{3}\right)^2 \cdot \frac{x}{2}\,dx = \frac{2}{9}$$

$$= \int_0^2 \left(x^2 - \frac{8}{3}x + \frac{16}{9}\right)\frac{x}{2}\,dx$$

$$= \frac{1}{2}\left[\frac{x^4}{4} - \frac{8}{3}\frac{x^3}{3} + \frac{16}{9}\frac{x^2}{2}\right]_0^2$$

$$= \frac{1}{2}\left[4 - \frac{64}{9} + \frac{32}{9}\right] = \frac{2}{9}.$$

(iii) standard deviation $= \sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$

(iv) $E(3X^2 - 2X) = \int_{-\infty}^{\infty}(3x^2-2x)\,f(x)\,dx$

$$= \int_0^2 (3x^2-2x)\left(\frac{x}{2}\right)dx = \frac{10}{3}$$

Th: $E(k) = k$

Th: $E(kX) = kE(X)$

Measures of Central Tendency:

(i) Mean
(ii) median
(iii) Mode

(i) **Mean (Arithmetic mean)**: $\bar{x} = \frac{\Sigma x_i}{n}$

(ii) **Median**: The median of a series of data is defined as that value which divides the whole series in to two equal parts

(iii) **Mode**: The mode is the value of a variate that occurs most often. i.e. the point having maximum frequency.

Moments :

If X be a random variable then r^{th} moment about any point a is given by

$$\mu_r' = E[(x-a)^r] = \begin{cases} \sum_i p_i (x_i-a)^r ; & X \text{ is discrete R.V.} \\ \int_{-\infty}^{\infty} f(x)(x-a)^r dx ; & X \text{ is conti. R.V.} \end{cases}$$

(a) Moments about origin :

$$\mu_r' = E(x^r) = \begin{cases} \sum_i x_i^r p_i ; & X \text{ is discrete R.V.} \\ \int_{-\infty}^{\infty} x^r f(x) dx ; & X \text{ is conti. R.V.} \end{cases}$$

(b) Moments about mean or Central Moment :

$$\mu_r = E[(X-\bar{x})^r] = \begin{cases} \sum_i (x_i-\bar{x})^r p_i ; & X \text{ is discrete R.V.} \\ \int_{-\infty}^{\infty} (x-\bar{x})^r f(x) dx ; & X \text{ is conti. R.V.} \end{cases}$$

Remark (i) $\mu_0 = 1$
$\mu_0' = 1$

(ii) First moment $\mu_1 = \sum p_i x_i - \bar{x}\sum p_i = \bar{x} - \bar{x} = 0$

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

(iii) Second moment $\boxed{\mu_2 = \sum (x_i-\bar{x})^2 p_i = \sigma^2 = \text{Variance}(X)}$

$$\mu_2 = \sum x_i^2 p_i - 2\sum p_i x_i \bar{x} + \sum p_i \bar{x}^2$$
$$= \mu_2' - 2\bar{x}.\bar{x} + \bar{x}^2$$
$$= \mu_2' - 2\mu_1'\mu_1' + \mu_1'^2$$

$$\boxed{\mu_2 = \mu_2' - \mu_1'^2}$$

(iii) Third moment $\mu_3 = \sum (x_i-\bar{x})^3 p_i$

$$\boxed{\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3}$$

(iv) Fourth moment $\boxed{\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4}$

Note: If instead of probability mass function we are given the corresponding frequency distribution then moment about any point is given by

$$\mu_r' = \frac{\sum_i (x_i - a)^r f_i}{\sum_i f_i}$$

Q The first four moments of a distribution about the value 5 are $-4, 22, -117$ and 560, obtain the moment about (i) mean and (ii) origin

Sol^n: Moments about 5 are given

i.e. $\mu_1'' = [E(X-5)^1] = -4$, ~~$\mu_1' = \bar{x} \Rightarrow -4 = \bar{x}$~~

let Moments about 5 are $\mu_1'', \mu_2'', \mu_3'', \mu_4''$

$\mu_2'' = [E(X-5)^2] = 22$

$\mu_3'' = [E(X-5)^3] = -117$

$\mu_4'' = [E(X-5)^4] = 560$

$\because \mu_1'' = [E(X-5)] = \sum (x_i - 5) p_i$ ($\because$ Moment about 5)

$-4 = \sum x_i p_i - 5 \sum p_i$

$\left[\because \sum x_i p_i = E(X) = \bar{X} \text{ and } \sum p_i = 1 \right]$

$-4 = E(X) - 5$

$\Rightarrow -4 = \bar{x} - 5$

or $\boxed{\bar{x} = 1}$ — (1)

Moment about mean i.e. ($\bar{x} = 1$)

$\mu_r = [E(X-\bar{x})^r] = \sum_i (x_i - \bar{x})^r p_i$

$\boxed{\mu_1 = 0}$

$\because \mu_2 = \mu_2' - \mu_1'^2$ or $\mu_2 = \mu_2'' - \mu_1''^2$

$\boxed{\mu_2 = 22 - (-4)^2 = 22 - 16 = 6}$

$\because \mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$ or $\mu_3 = \mu_3'' - 3\mu_1''\mu_2'' + 2\mu_1''^3$

$\mu_3 = (-117) - 3(-4)(22) + 2(-4)^3$

$= -117 + 264 - 128$

$\boxed{\mu_3 = 19}$

8 ʀ3 (B)

$\because \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$ or $\mu_4 = \mu_4'' - 4\mu_3''\mu_1'' + 6\mu_2''\mu_1''^2 - 3\mu_1''^4$

$\mu_4 = 560 - 4(-117)(-4) + 6(22)(-4)^2 - 3(-4)^4$

$\boxed{\mu_4 = 32}$

Moment about origin:

$\because \boxed{\mu_1' = E[X] = \bar{X} = 1}$ (from ①.)

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

$\mu_2' = \mu_2 + \mu_1'^2$

$\boxed{\mu_2' = 6+1 = 7}$

$\because \mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$

or $\mu_3' = \mu_3 + 3\mu_1'\mu_2' - 2\mu_1'^3$

$\mu_3' = 19 + 3(1)(7) - 2(1)^3$

$\boxed{\mu_3' = 38}$

$\because \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

or $\mu_4' = \mu_4 + 4\mu_3'\mu_1' - 6\mu_2'\mu_1'^2 + 3\mu_1'^4$

$\mu_4' = 32 + 4(38)(1) - 6(7)(1) + 3(1).$

$\boxed{\mu_4' = 145}$

Moment Generating function: (mgf)

The moment generating function (mgf) of a random variable X ~~having the probability function f(x), is~~ is given by.

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} P(x) & \text{for discrete RV with pmf } P(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{for continuous RV with pdf } f(x) \end{cases}$$

provided the right hand side is absolutely convergent for some positive number h such that $-h<t<h$ where t is any real parameter.

Now $M_X(t) = E(e^{tX}) = E\left[1 + tX + \frac{t^2 X^2}{2!} + \cdots + \frac{t^r X^r}{r!} + \cdots\right]$

$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots + \frac{t^r}{r!}E(X^r)$

$= 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \cdots + \mu_r' \frac{t^r}{r!} + \cdots$

where μ_r' is the rth order moment about origin.

Since $M_X(t)$ generates moments, hence it is known as moment generating function.

Also $\mu_r' = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}$

Properties of Moment Generating function :

(i) Moment generating function about mean $\bar{X}$

$M_X(t) = E\left[e^{t(X-\bar{X})}\right] = E\left[1 + t(X-\bar{X}) + \frac{t^2}{2!}(X-\bar{X})^2 + \cdots\right]$

$= E(1) + tE(X-\bar{X}) + \frac{t^2}{2!}E\{(X-\bar{X})^2\} + \cdots$

$M_X(t) = 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \cdots + \mu_r \frac{t^r}{r!} + \cdots$

or $\boxed{\mu_r = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}}$

(ii) ~~mgf~~ M.G.F about any pt : $(x = a)$

$M_X(t) = E\left[e^{t(X-a)}\right] = 1 + t\mu_1'' + \frac{t^2}{2!}\mu_2'' + \cdots + \frac{t^r}{r!}\mu_r'' + \cdots$

(iii) If X and Y are two independent R.V.

then $M_{X+Y}(t) = M_X(t) M_Y(t)$

(iv) If $X_2 = c_1 X_1 + c_2$, then

$M_{X_2}(t) = e^{tc_2} M_{X_1}(c_1 x_1)$

R. 317.
R.V

Q Let the random variable X assume the value 'r' with the probability law $P(X=r) = q^{r-1}p$, $r = 1,2,3,\ldots$ Find the mgf of X and hence its mean and variance. Verify it by finding the mean from usual definition.

Solⁿ:

$$\text{Mgf of } X = M_X(t) = E(e^{tX}) = \sum_{r=1}^{\infty} e^{tr} p_r$$

∵ We know that mean $= \bar{x} = \mu_1'$ = first moment about origin

$$= \sum_{r=1}^{\infty} e^{tr} q^{r-1} p$$

$$= \frac{p}{q} \sum_{r=1}^{\infty} (q e^t)^r$$

$$= \frac{p}{q}\left[q e^t + (q e^t)^2 + (q e^t)^3 + \ldots\right]$$

$$= \frac{p}{q}\, q e^t \left[1 + q e^t + (q e^t)^2 + \ldots\right]$$

$$M_X(t) = p e^t \frac{1}{(1-qe^t)} = \frac{p e^t}{(1-qe^t)}$$

Now Mean $= \mu_1' = \left[\frac{d\, M_X(t)}{dt}\right]_{t=0}$

$$= p\left[\frac{(1-qe^t)e^t - e^t(-qe^t)}{(1-qe^t)^2}\right]_{t=0} = \left[\frac{pe^t}{(1-qe^t)^2}\right]_{t=0}$$

$$\mu_1' = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

[∵ $p+q=1$

and $\mu_2' = \left[\frac{d^2 M_X(t)}{dt^2}\right]_{t=0} = p\left[\frac{(1-e^tq)^2 e^t - 2e^t(1-qe^t)(-qe^t)}{(1-qe^t)^4}\right]_{t=0}$

$$= p\left[\frac{(1-e^tq)\, e^t(1-qe^t+2qe^t)}{(1-qe^t)^4}\right]_{t=0}$$

$$= \left[pe^t \frac{(1+qe^t)}{(1-qe^t)^3}\right]_{t=0} = \frac{p(1+q)}{(1-q)^3} = \frac{p(1+q)}{p^3} = \frac{1+q}{p^2}$$

$\bar{x}$ = mean $= \mu_1'$ = first moment about origin

Var $= \sigma^2 = \mu_2$ = second moment about mean

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

$$\mu_2 = \sigma^2 = \frac{1+q}{p^3} - \frac{1}{p^2} = \frac{q}{p^2}$$

By usual definition

$$E(X) = \bar{X} = \text{mean} = \sum_{r=1}^{\infty} r P(X=r) = \sum_{r=1}^{\infty} r(q^{r-1}p)$$

$$\bar{x} = p[1 + 2q + 3q^2 + \dots]$$
$$= p(1-q)^{-2}$$
$$= \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Q Find Mean and Standard deviation of the exponential distribution.

Sol^n: If X is a continuous random variable and exponentially distributed having the following p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} ; & 0 < x < \infty. \\ 0 ; & \text{otherwise} \end{cases}$$

$\because$ mean $= \bar{x} = \mu_1'$ = moment about origin

where λ is any parameter

$M_X(t) = E[e^{xt}]$

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

Now, M.G.F. of exponential distribution

$$M_X(t) = \int_{-\infty}^{\infty} \lambda e^{-\lambda x} e^{tx} dx$$
$$= \lambda \int_0^{\infty} e^{-(t+\lambda)x} dx$$
$$= \lambda \left\{ \frac{e^{-(\lambda - t)}}{-(\lambda - t)} \right\}_0^{\infty}$$

$$\therefore M_X(t) = \frac{\lambda}{(\lambda - t)}$$

Or $M_X(t) = \frac{1}{(1-\frac{t}{\lambda})} = \left(1 - \frac{t}{\lambda}\right)^{-1}$

$$M_X(t) = 1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \dots \quad \text{—— (1)}$$

By the definition of M.G.F we have.

$$M_X(t) = 1 + \mu_1'\{t\} + \mu_2'\left\{\frac{t^2}{2!}\right\} + \mu_3' \frac{t^3}{3!} + \dots \quad \text{—— (2)}$$

From (1) and (2),

$\boxed{\mu_1' = \frac{1}{\lambda}}$ mean $= \bar{x}$

$\mu_2' = \frac{2!}{\lambda^2}$

$\mu_3' = \frac{3!}{\lambda^3}$

Variance $= \mu_2 = \sigma^2 = \mu_2' - \mu_1'^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

S.D $= \sigma = \frac{1}{\lambda}$

23 R.
R.V

Correlation Coefficient: Karl Pearson defined the four coefficients based on central moments

(i) β coefficients:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \because\ \mu_2 = \text{Vari}(X) = \sigma^2$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \text{ (Called measure of Kurtosis)}$$

(ii) γ coefficients:

$$\gamma_1 = \pm\sqrt{\beta_1} \quad \text{or} \quad \gamma_1 = \frac{\mu_3}{\sigma^3} \text{ (Called Coefficient of skewness)}$$

$$\gamma_2 = \beta_2 - 3 \text{ (Called Coefficient of Kurtosis)}$$

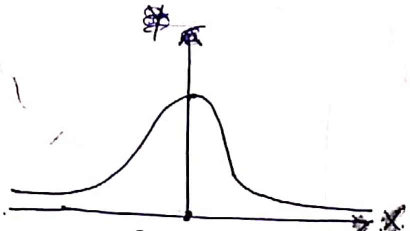
Karl Pearson's coefficient of skewness denoted by S_K, is given by
$S_K = \frac{\text{Mean} - \text{Mode}}{\sigma}$

* If Mean = Mode, then $S_K = 0$, symm. distri.
* $S_K > 0$, +ve skewed distri.
* $S_K < 0$, -ve skewed distri.

Skewness: Skewness is the measure of the shape of the curve and not of its size. It is the deviation from symmetry. (Skewness of X is the third moment of the standard score of X i.e. $\text{skew}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$)

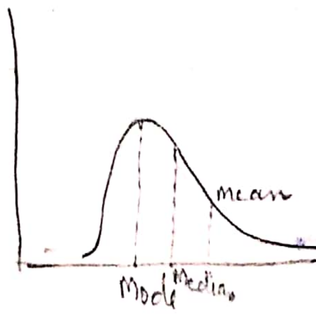
Symmetrical distribution: Mean = Mode = Median

(when $\beta_1 = 0$, then the distribution of X is said to be unskewed)

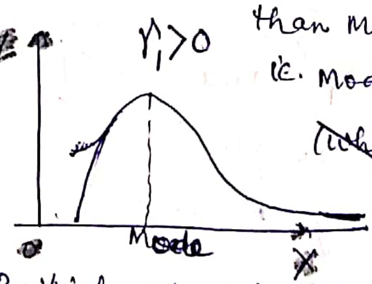


Symmetrical distribution

Positively skewed distribution: Mean is greater than Mode or median
i.e. Mode < median < Mean



$\gamma_1 > 0$

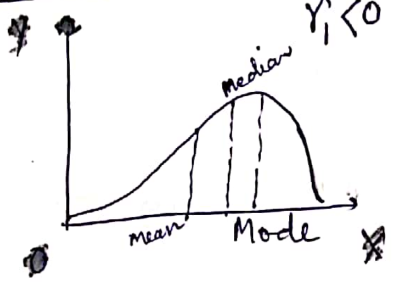


Positively skewed dist.

~~(when $\beta_1 > 0$, then the distribution of X is said be positively skewed)~~

If the distribution is positively skewed then probability density function has a long tail to the right.

Negatively Skewed Distribution:



$\gamma_1 < 0$. Mean is less than mode and median

ie. Mean < Median < Mode

~~(when $\beta_1 < 0$, then the distribution of X is said to be negatively skewed)~~

Negatively skewed distribution

If the distribution is negatively skewed then the probability density fn has a long tail to the left.

Note: (i) Empirical relationship Mode = 3Median − 2Mean

(ii) If $\beta_1 = 0$, the curve is symmetrical.

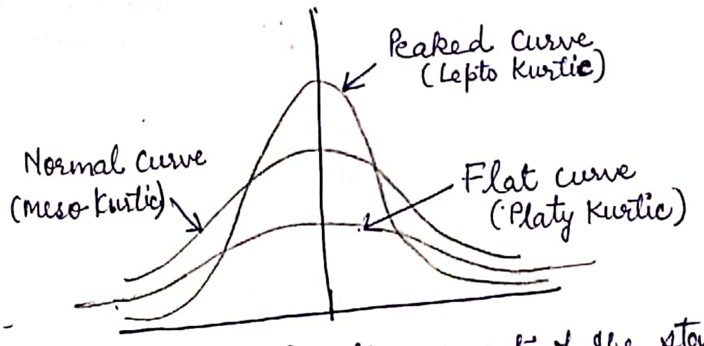
Hence β_1 can be taken as measure of skewness.

Kurtosis: The flatness of the mode is called Kurtosis. β_2 is taken as the measure of Kurtosis.

(i) If $\beta_2 = 3$, then $\gamma_2 = 0$, curve is called meso-kurtic.

(ii) If $\beta_2 > 3$, $\gamma_2 > 0$, curve is called leptokurtic.

(iii) If $\beta_2 < 3$, $\gamma_2 < 0$, curve is called platy-kurtic.



Note: Kurtosis of X is the fourth moment of the standard score

$$\text{Kurt}(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

Q. Calculate the first four moments about mean for the following distribution and also hence β_1 and β_2.

x:	0	1	2	3	4	5	6	7	8
y:	1	8	28	56	70	56	28	8	1

Soln: Here mean $= \dfrac{\Sigma fx}{\Sigma f} = \dfrac{1024}{256} = 4$

x_i	f_i	(x_i-4)	$f_i(x_i-4)$	$f_i(x_i-4)^2$	$f_i(x_i-4)^3$	$f_i(x_i-4)^4$
0	1	−4	−4	16	−64	256
1	8	−3	−24	72	−216	648
2	28	−2	−56	112	−224	448
3	56	−1	−56	56	−56	56
4	70	0	0	0	0	0
5	56	1	56	56	56	56
6	28	2	56	112	224	448
7	8	3	24	72	216	648
8	1	4	4	16	64	256
Total 36	256	0	0	512	0	2816

Hence moments about mean $x=4$ are

$$\mu_1 = \frac{\Sigma f_i(x_i-4)}{\Sigma f_i} = 0 \quad ; \quad \mu_2 = \frac{\Sigma f_i(x_i-4)^2}{\Sigma f_i} = \frac{512}{256} = 2$$

$$\mu_3 = \frac{\Sigma f_i(x_i-4)^3}{\Sigma f_i} = 0 \quad ; \quad \mu_4 = \frac{\Sigma f_i(x_i-4)^4}{\Sigma f_i} = \frac{2816}{256} = 11$$

Also $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = 0$; $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{11}{4} = 2.75$.

As $\beta_1 = 0$, hence curve is symmetric about mean and $\beta_2 < 3$ hence curve is platykurtic in nature.

Q For a distribution mean is 10, variance is 16, γ_1 is 1 and β_2 is 4. Obtain the first four moments about origin. Also comment upon the nature of distribution.

Soln: Given $\bar{x} = \mu_1' = 10$

$\mu_2 = 16$

$\gamma_1 = 1$

$\beta_2 = 4$

Now $\gamma_1 = \sqrt{\beta_1}$

$1 = \sqrt{\beta_1}$ or $\beta_1 = 1$ — (1)

$\because \beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

or $\beta_1^2 = \dfrac{\mu_3}{\mu_2^{3/2}}$

$\Rightarrow 1 = \dfrac{\mu_3}{(16)^{3/2}}$ or $\mu_3 = (16)^{3/2} = 64$ — (2)

and $\beta_2 = \frac{\mu_4}{\mu_2^2} \Rightarrow 4 = \frac{\mu_4}{(16)^2}$

or $\mu_4 = 1024$ — (3)

Hence, we get

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

$\therefore \mu_2' = \mu_2 + \mu_1'^2$

$\boxed{\mu_2' = 16 + (10)^2 = 116}$ — (4)

$\because \mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$

or $\mu_3' = \mu_3 + 3\mu_1'\mu_2' - 2\mu_1'^3$

$\mu_3' = 64 + 3(10)(116) - 2(10)^3$

$= 64 + 3480 - 2000$

$\boxed{\mu_3' = 1544}$ — (5)

Similarly.

$\because \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$\mu_4' = \mu_4 + 4\mu_3'\mu_1' - 6\mu_2'\mu_1'^2 + 3\mu_1'^4$

$\mu_4' = 1024 + 4(1544)(10) - 6(116)(10)^2 + 3(10)^4$

$= 1024 + 61760 - 69600 + 30000$

$\boxed{\mu_4' = 23184}$ — (6)

Nature of distribution

Here $\beta_1 = 1 \neq 0$ Hence distribution is not symmetric
and $\beta_2 = 4 > 3$ Hence curve is peaked curve i.e. distribution is lepto kurtic in nature.

~~Chebyst~~ Chebyshev's inequality: If X is random variable with mean $\bar{x}$ and variance σ^2, then

$P[|X-\bar{x}| \geq \lambda] \leq \frac{\sigma^2}{\lambda^2}$ ~~any~~ where $\lambda > 0$

or $P[|X-\bar{x}| < \lambda] \geq 1 - \frac{\sigma^2}{\lambda^2}$

Proof: Let X be a continuous random variable having p.d.f. f(x), then by the definition of variance, we have

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

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

$$\sigma^2 \geq \int_{-\infty}^{\bar{x}-\lambda} (x-\bar{x})^2 f(x)\,dx + \int_{\bar{x}+\lambda}^{\infty} (x-\bar{x})^2 f(x)\,dx \quad \text{——— (1)}$$

in the first integral

$$x \leq (\bar{x}-\lambda)$$

$$\text{or } (x-\bar{x})^2 \geq \lambda^2$$

in the second integral

$$x \geq (\bar{x}+\lambda)$$

$$\text{or } (x-\bar{x})^2 \geq \lambda^2$$

using these results in (1), we have

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

$$\sigma^2 \geq \lambda^2\left[\int_{-\infty}^{\bar{x}-\lambda} f(x)\,dx + \int_{\bar{x}+\lambda}^{\infty} f(x)\,dx\right]$$

$$\sigma^2 \geq \lambda^2\left[P(X \leq \bar{x}-\lambda) + P(X \geq \bar{x}+\lambda)\right]$$

$$\sigma^2 \geq \lambda^2\left[P(\bar{x}+\lambda \leq X \leq \bar{x}-\lambda)\right]$$

$$\text{or } \frac{\sigma^2}{\lambda^2} \geq P\left[|X-\bar{x}| \geq \lambda\right]$$

$$\text{or } P\left[|X-\bar{x}| < \lambda\right] \geq 1 - \frac{\sigma^2}{\lambda^2}, \quad \lambda > 0$$

OR

$$\frac{\sigma^2}{\lambda^2} \geq \left[P(X \leq \bar{x}-\lambda) + \{1 - P(X \leq \bar{x}+\lambda)\}\right]$$

$$\text{or } -\frac{\sigma^2}{\lambda^2} \leq \left[P(X \leq \bar{x}+\lambda) - P(X \leq \bar{x}-\lambda) - 1\right]$$

$$\text{or } 1 - \frac{\sigma^2}{\lambda^2} \leq P\left[(\bar{x}-\lambda) \leq X \leq (\bar{x}+\lambda)\right]$$

$$\text{or } 1 - \frac{\sigma^2}{\lambda^2} \leq P\left[|X-\bar{x}| < \lambda\right], \quad \lambda > 0$$

Proved.

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

Solⁿ: Using chebyshev's inequality

$$P\{|X-\bar{x}| \geq \lambda\} \leq \frac{\sigma^2}{\lambda^2}, \quad \lambda > 0$$

$$\text{or } P\{|X-\bar{x}| < \lambda\} \geq 1 - \frac{\sigma^2}{\lambda^2}$$

$$\text{or } P\{\bar{x}-\lambda \leq X \leq \bar{x}+\lambda\} \geq 1 - \frac{\sigma^2}{\lambda^2} \quad ; \bar{x} = 12, \sigma^2 = 9$$

$$P\{12-\lambda \leq X \leq 12+\lambda\} \geq 1 - \frac{9}{\lambda^2}$$

Let $\lambda = 6$

$\Rightarrow P\{6 < X < 18\} \geq 1 - \frac{9}{36}$

$\geq \frac{27}{36}$

$\geq \frac{3}{4}$

Q. A random variable X has a mean 10 and a variance 4 and unknown probability distribution. Find the value of c such that $P\{|X-10| \geq c\} \leq 0.04$.

Soln $c = 10$ $\because$ Chebyshev's inequality

$$P\{|X-\bar{X}| \geq \lambda\} \leq \frac{\sigma^2}{\lambda^2}$$

Given that $P\{|X-10| \geq c\} \leq 0.04$

$\Rightarrow$ $\bar{x} = 10$

$\lambda = c$

$\frac{\sigma^2}{\lambda^2} = 0.04$ (It is given in the question that $\sigma^2 = \text{Var.} = 4$)

$\Rightarrow \frac{4}{\lambda^2} = 0.04$

or $\lambda = 10$

$\Rightarrow \boxed{\lambda = c = 10}$

Q Two dice are thrown once. If X is the sum of the numbers showing up, prove that $P\{|X-7|\geq 3\} \leq \frac{35}{54}$. Compare this value with the exact probability.

Solⁿ:

X :	2	3	4	5	6	7	8	9	10	11	12
p_i :	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$\bar{x}$ = mean = $E(X) = \Sigma x_i p_i = 7$

$\sigma^2 = Var.(X) = E\{(X-\bar{x})^2\} = E(X^2) - \{E(X)\}^2$

$\sigma^2 = \Sigma x_i^2 p_i - (\bar{x})^2$

$\sigma^2 = \frac{1974}{36} - 49 = \frac{1974-1764}{36}$

$\sigma^2 = \frac{210}{36} = \frac{35}{6}$

By Chebyshev's inequality

$P\{|X-\bar{x}| \geq \lambda\} \leq \frac{\sigma^2}{\lambda^2}$

Comparing with $P\{|X-7| \geq 3\}$

$\bar{x} = 7, \quad \lambda = 3$

then $\frac{\sigma^2}{\lambda^2} = \frac{35}{6} \times \frac{1}{9} = \frac{35}{54} = 0.6481$

$\Rightarrow \boxed{P\{|X-7| \geq 3\} \leq 0.6481}$

Actual probability is given by.

$P\{|X-7| \geq 3\} = P\{7+3 \leq X \leq -3+7\}$

$= P\{10 \leq X \leq 4\}$

$= P\{X = 2, 3, 4, 10, 11, 12\}$

$= \frac{1}{3} = 0.33$

Q. A random variable X is exponentially distributed with parameter 1. Use chebyshev's inequality to show that $P\{-1 \le X \le 3\} \ge \frac{3}{4}$. Find the actual probability also.

Solⁿ: $\because$ For exponentially distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0<x<\infty \\ 0, & \text{otherwise} \end{cases}$$

Here given that parameter $\lambda = 1$

$\Rightarrow$ pdf is given by $f(x) = \begin{cases} e^{-x}, & 0<x<\infty \\ 0, & \text{otherwise} \end{cases}$

Now $\bar{x} = \text{mean} = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

$$\bar{x} = \int_0^{\infty} x e^{-x}\,dx = 1$$

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

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

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

$$\sigma^2 = 2 - 1 = 1$$

By chebyshev's inequality $\quad P\{|X-\bar{x}| < \lambda\} \ge 1 - \frac{\sigma^2}{\lambda^2}$

$\cancel{P\{|X-\bar{x}| \ge \lambda\} \le \frac{\sigma^2}{\lambda^2}}$

or $P\{(\bar{x}-\lambda) < X < (\bar{x}+\lambda)\} \ge 1 - \frac{\sigma^2}{\lambda^2}$

or $\cancel{P\{(\bar{x}+\lambda) \le X \le (\bar{x}-\lambda)\} \le \frac{\sigma^2}{\lambda^2}}$

or $P\{(1-\lambda) < X < (1+\lambda)\} \ge 1 - \frac{\sigma^2}{\lambda^2}$

$\cancel{P\{(1+\lambda) \le X \le (1-\lambda)\} \le \frac{\sigma^2}{\lambda^2}}$

Comparing with $P\{-1 \le X \le 3\}$,

we have $\lambda = 2$

$$\Rightarrow P\{-1 \le X \le 3\} \ge 1 - \frac{1}{4}$$

$$\text{or } P\{-1 \le X \le 3\} \ge \frac{3}{4} = 0.75$$

The actual probability is given by

$$P\{-1 \le X \le 3\} = \int_{-1}^{3} f(x)\,dx = \int_0^3 e^{-x}\,dx = 1 - e^{-3}$$

$$= 0.9502$$

Normal distribution:

The most important continuous probability distribution used in statistics is normal distribution. It is a limiting form of the binomial distribution, in which p is not small but $n \to \infty$.

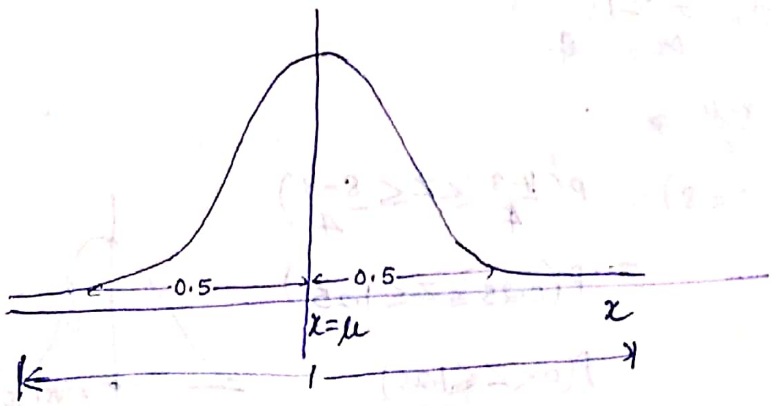
Definition: A random variable X is said to have a normal distribution with parameters μ(mean) and σ^2(variance) if its probability density function is given by

$$\cancel{f(x)=}\ p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$
$$-\infty < \mu < \infty, \sigma > 0$$

Note: 1) The normal distribution with mean μ and variance σ^2 can be denoted by the symbol $N(\mu, \sigma^2)$.

2) The probability between two specified values a & b is

$P(a<x<b) =$ Area under the curve $p(x)$ between the specified values $X=a$ & $X=b$.



Normal probability curve

The curve is bell-shaped and symmetrical about the line $x=\mu$.

Mean, median and mode of the distribution coincide

3) The normal distribution is often called Gaussian distribution.

4) Some of the important continuous distributions are Uniform distribution, Gamma distribution, Exponential and Normal distributions.

8.4.7

Standard form of the normal distribution :

The probability density function for the normal distribution in standard form is given by

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

where $z = \frac{x-\mu}{\sigma}$, z is called the standard normal random variable

Note (i) standard form of the normal distribution is free from any parameter.

(ii) for standard normal variates

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

and $$P(z \le 0) = P(z \ge 0) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{z^2}{2}} dz = \frac{1}{2}$$

Q If X is $N(3,16)$, then find $P(4 \le X \le 8)$

Solⁿ: $\mu = 3$, $\sigma^2 = 16$
or $\sigma = 4$

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

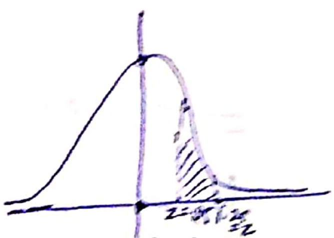
$$P(4 \le X \le 8) = P\left(\frac{4-3}{4} \le Z \le \frac{8-3}{4}\right)$$

$$= P(0.25 \le Z \le 1.25)$$

$$= P(0 < Z \le 1.25) - P(0 < Z \le 0.25)$$

$$= 0.3944 - 0.0987$$

$$= 0.2957$$



Note :: In conti. R.V. prob. of x lying in the small interval $(x - \frac{dx}{2}, x + \frac{dx}{2})$ is $f(x)dx$ i.e. $P\{x - \frac{dx}{2} < x < x + \frac{dx}{2}\} = f(x)dx$

Now in Normal dist. $f(x)dx = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$; for standard normal distr. take $\frac{x-\mu}{\sigma} = z$, $\frac{dx}{\sigma} = dz$

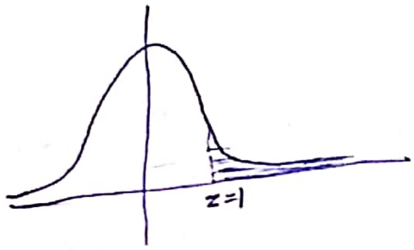
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = f(z)\, dz$$

L.4.3
Mathematical Exp. and Theo. distri.

Q The distribution of weekly wages for 500 workers in a factory is approximately normal with the mean and standard deviation of Rs. 75 and Rs. 15. Find the number of workers who receive weekly wages:
(i) more than Rs 90. (ii) less than Rs. 45.

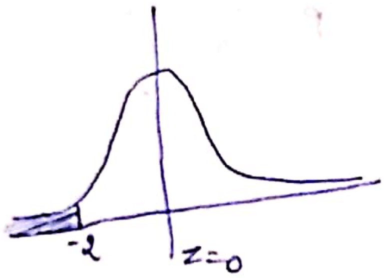
Soln: Given $N = 500$, $\mu = 75$, $\sigma = 15$ $\qquad \because Z = \frac{X-\mu}{\sigma}$

(i) $P(X>90) = P\left(Z > \frac{90-75}{15}\right)$
$= P(Z > 1)$
$= 0.5 - P(0<Z<1)$
$= 0.5 - 0.3413$
$= 0.1587$



∴ No. of workers receiving weekly wages more than 90 Rs $= 500 \times 0.1587$
$= 79.35 \cong 79$

(ii) $P(X<45) = P\left(Z < \frac{45-75}{15}\right)$
$= P(Z < -2)$
$= 0.5 - P(0<Z<2)$
$= 0.5 - 0.4772$
$= 0.0228$



No. of workers receiving weekly wages less than 45 Rs $= 500 \times 0.0228$
$= 11.4 \cong 11$

Q For $-\infty < x < \infty$, and probability density
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$
Show that the total probability is 1

Proof: Total Probability is given by
$$P(-\infty < x < \infty) = \int_{-\infty}^{\infty} f_X(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left\{\frac{(x-\mu)}{\sqrt{2}\sigma}\right\}^2} dx$$

Scanned by CamScanner

taking $\frac{x-\mu}{\sqrt{2}\sigma} = z$

$dx = \sqrt{2}\,\sigma\, dz$

$$\Rightarrow \quad P = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} dz$$

$$P = \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} e^{-z^2} dz$$

let $z^2 = u$

$2z\,dz = du$

$$P = \frac{2}{\sqrt{\pi}} \int_0^{\infty} e^{-u} \frac{du}{2\sqrt{u}}$$

$$= \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-u}\, u^{-1/2} du$$

$$= \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-u}\, u^{\frac{1}{2}-1} du$$

$$\left\{ \begin{array}{l} \int_0^{\infty} e^{-u} u^{\frac{1}{2}-1} = \Gamma(\tfrac{1}{2}) \\ \text{and } \Gamma(\tfrac{1}{2}) = \sqrt{\pi} \end{array} \right.$$

$$P = \frac{1}{\sqrt{\pi}} \Gamma(\tfrac{1}{2}) = 1.$$

Cumulative distribution function (cdf) or simply distribution function:

If X is a continuous random variable, then $F(x) = P(X \le x)$ is called cdf that is $F(x) = P(X \le x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(x)\,dx$

where $f(x)$ is probability density function

The cumulative distribution function $F(x)$ has the following important properties.

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

(ii) $F(x)$ is a non-decreasing function, that is if $x_1 < x_2$, then $F(x_1) \le F(x_2)$

(iii) $F(-\infty) = 0$ & $F(\infty) = 1$

(iv) $f(x) = F'(x)$ at all pt where $F(x)$ is differentiable

Mean and Variance: If X is a continuous random variable and $f(x)$ is the pdf of X, then we define

$$\text{Mean} = \mu = \int_R x f(x)\,dx = E(X)$$

If $R = [a, b]$

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

$$\text{Variance} = \sigma^2 = \int_R (x-\mu)^2 f(x)\,dx = E\{(X-\mu)^2\}$$

$$\sigma^2 = \int_R x^2 f(x)\,dx - \mu^2$$

If $R = [a, b]$

then $\text{Variance} = \sigma^2 = \int_a^b x^2 f(x)\,dx - \mu^2 = E(X^2) - \{E(X)\}^2$

eg Find the mean and variance of the random variable X whose density function, f is defined by

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ 4x(1-x^2), & \text{if } 0 \le x \le 1 \\ 0 & \text{if } x > 1 \end{cases}$$

Soln

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 4x^2(1-x^2)\,dx = \frac{8}{15}$$

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

$$\sigma^2 = \int_0^1 4x^3(1-x^2)\,dx - \frac{64}{225}$$

$$\sigma^2 = \frac{1}{3} - \frac{64}{225} = \frac{11}{225}$$

eg2 X is a continuous random variable with p.d.f given by

$$f(x) = \begin{cases} 2x^3, & 0 \le x \le 1 \\ 2(2-x)^3, & 1 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

find the standard deviation and mean for the random variable X.

Sol^n:

$$\text{Mean} = \mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x(2x^3)dx + \int_1^2 2x(2-x)^3\,dx$$

$$= 1$$

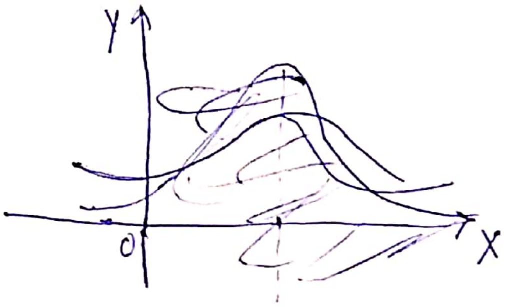
$$\text{Variance} = \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

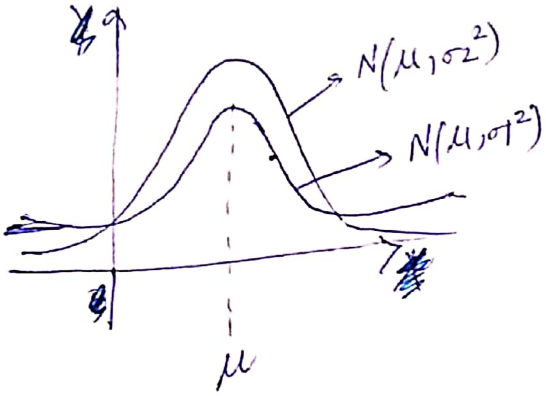
$$\sigma^2 = \left[\int_0^1 x^2(2x^3)\,dx + \int_1^2 2x^2(2-x)^3\,dx\right] - 1$$

$$\sigma^2 = \frac{16}{15} - 1 = \frac{1}{15}$$

$$\text{Standard deviation} = \sqrt{\text{Variance}}$$

$$\sigma = \sqrt{\frac{1}{15}} = 0.258$$

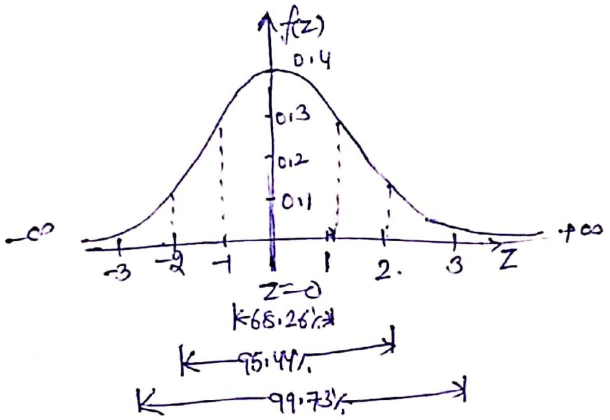




$\sigma_1 > \sigma_2$

Normal probability curve $\sigma_1 > \sigma_2$.

Standardised normal curve



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

$x = \mu - \sigma$ to $x = \mu + \sigma$

i.e. $z = -1$ to $z = 1$

* $P(\mu - \sigma < x < \mu + \sigma)$

$= P(-1 < z < 1) = 68.26\%$

* $P(\mu - 2\sigma < x < \mu + 2\sigma)$

$= P(-2 < z < 2) = 95.44\%$

* $P(\mu - 3\sigma < x < \mu + 3\sigma)$

$= P(-3 < z < 3) = 99.73\%$

Fitting of Normal distribution: In order to fit a normal distribution to a given frequency distribution x_i and f_i, $i = 1, 2, \ldots, n$.

we find $\mu = \frac{\sum f_i x_i}{\sum f_i}$ and $\sigma^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i}\right)^2$

from the given data. Hence the normal curve fitted to the given data is given by.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \; ; \; -\infty < x < \infty$$

Q Fit a normal curve to the following frequency distribution :

x :	4	6	8	10	12	14	16	18	20	22	24
y :	1	7	15	22	35	43	38	20	13	5	1

Solⁿ we form the following table

x	f = y	fx	fx²
4	1	4	16
6	7	42	252
8	15	120	960
10	22	220	2200
12	35	420	5040
14	43	602	8428
16	38	608	9728
18	20	360	6480
20	13	260	5200
22	5	110	2420
24	1	24	576
	$\Sigma f = 200$	$\Sigma = 2770$	$\Sigma = 41300$

$\because \sigma^2 = \mu_2 = \mu_2' - \mu_1'^2$

$\therefore \mu = \cancel{\Sigma fx} \frac{\Sigma fx}{\Sigma f} = \frac{2770}{200} = 13.85 = \text{mean} = E(X) = \mu_1'$

$\sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2 = \frac{41300}{200} - \left(\frac{2770}{200}\right)^2 = 206.5 - 191.8225$
$= 14.6775$

$\sigma = 3.83$

Hence the normal curve to be fitted is :

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

with $\mu = 13.85$ and $\sigma = 3.83$

Lecture - 3

Poisson's Distribution : Poisson distribution is a limiting case of the binomial distribution under the following conditions.

1.) n, the number of trials is ~~iden~~ indefinitely large, i.e. $n \to \infty$
2.) p, the probability of success for each trial is indefinitely small, i.e. $p \to 0$
3.) $np = \lambda$ (say) is finite positive real number.

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

The probability of r success in a series of n independent trials is

$$P(r) = {}^nC_r \, p^r q^{n-r}$$

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

$$= \frac{n(n-1)\cdots(n-r+1)}{r!} \left(\frac{\lambda}{n}\right)^r \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

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

$$\lim_{\substack{n \to \infty \\ p \to 0}} P(r) = \frac{\lambda^r e^{-\lambda}}{r!}, \quad r = 0, 1, \dots$$

$$\left\{ \because \lim_{n \to \infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda} \right\}.$$

This limiting form of Binomial distribution with above probability is called Poisson's distribution.

Note 1.) λ is known as the parameter of the distribution.

2.) $e = 2.7183$

3.) $\displaystyle\sum_{r=0}^{\infty} P(X=r) = \sum_{r=0}^{\infty} \frac{e^{-\lambda}\lambda^r}{r!} = e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right] = e^{-\lambda} e^{\lambda} = 1$

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non negative values and its probability mass function is given by.

$$P(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x=0,1,2\ldots, \lambda>0 \\ 0, & \text{otherwise} \end{cases}$$

Note: This distribution is used to describe the behaviour of rare events such as the number of accidents on road, number of printing mistakes in a book etc.

Q. Suppose on an average 1 house in 1,000 in a certain district has a fire during a year. If there are 2,000 houses in that district, what is the probability that exactly 5 houses will have a fire during the year?

Soln: $n = 2000$, $p = \frac{1}{1000}$

$\lambda = np = 2000 \times \frac{1}{1000} = 2.$

Required probability that exactly 5 houses will have a fire during the year $= P(5)$

$$= \frac{e^{-\lambda}\lambda^5}{5!}$$

$$= \frac{e^{-2}2^5}{5!}$$

$$= \frac{.135 \times 32}{120}$$

$$= .036$$

Scanned by CamScanner

Mean and variance of the Poisson distribution:

for the Poisson distribution

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

$$E(X) = \text{Mean} = \mu = \sum_{r=0}^{\infty} r\, P(r)$$

$$= \sum_{r=0}^{\infty} r \frac{\lambda^r e^{-\lambda}}{r!}$$

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

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

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

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

$$= \lambda$$

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

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

$$= \sum_{r=0}^{\infty} r^2 P(X=r) - \lambda^2$$

$$= \sum_{r=0}^{\infty} r^2 \frac{\lambda^r e^{-\lambda}}{r!} - \lambda^2$$

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

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

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

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

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

$$= \lambda e^{-\lambda} \cdot e^{\lambda}(1+\lambda) - \lambda^2$$

$$= \lambda$$

Hence, standard deviation $\sigma = \sqrt{\text{var}(X)} = \sqrt{\lambda}$

Fitting a Poisson Distribution: When a Poisson distribution is to be fitted to observe data, the following procedure is adopted.

1.) Compute the mean $\bar{X}$ and take it equal to the mean of the fitted (Poisson) distribution.

$$\bar{X} = \lambda$$

2.) Obtain the probabilities. $P(X=r) = \frac{e^{-\lambda}\lambda^{r}}{r!}$, $r = 0, 1, 2 \ldots\ldots$

3.) The expected or theoretical frequencies according to Poisson distribution can be calculated as

$$f(r) = N \cdot P(X=r)$$

where N is the total observed frequency.

Q. Data was collected over a period of 10 years, showing number of deaths from horse kicks in each of the 200 army corps. The distribution of deaths was as follows.

No. of deaths :	0	1	2	3	4	Total
Frequency :	109	65	22	3	1	$200 = N = \Sigma f_i$

Fit a Poisson distribution to the data and calculation the theoretical frequencies.

Sol^n.

x	f	fx
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
	$\Sigma f = 200$	$\Sigma f x = 122$

$$\cancel{\bar{x} = \Sigma fx} \quad \bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{122}{200} = 0.61 = \lambda.$$

$\lambda = .3.5$

Math. Expec. and Theo. Distri.

X	$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Frequency $NP(X=x)$
0	$e^{-.61}\frac{(.61)^0}{0!} = .5432$	$200 \times .5432 = 108.64 \approx 109$
1	$e^{-.61}\frac{(.61)^1}{1!} = .3313$	$200 \times .3313 = 66.27 \approx 66$
2	$e^{-.61}\frac{(.61)^2}{2!} = .101$	$200 \times .101 = 20.21 \approx 20$
3	$e^{-.61}\frac{(.61)^3}{3!} = .021$	$200 \times .021 = 4.11 \approx 4$
4	$e^{-.61}\frac{(.61)^4}{4!} = .003$	$200 \times .003 = .63 \approx 1.$

Recurrence formula for the Poisson Distribution:

$$\because \; P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

$$P(r+1) = \frac{e^{-\lambda}\lambda^{r+1}}{(r+1)!}$$

$$\Rightarrow \; P(r+1) = \frac{\lambda}{(r+1)} \cdot P(r).$$

Q If the variance of the Poisson distribution is 2, find the probabilities for $r = 1, 2, 3, 4$ from the recurrence relation of the Poisson distribution.

Soln: Here $\lambda = 2$

$\because \; P(r+1) = \frac{\lambda}{(r+1)} P(r) = \frac{2}{(r+1)} P(r)$. which is the recurrence relation

$P(1) = 2 \cdot P(0) = 2 \cdot e^{-2} = 2 \times .1353 = .2706$ $\quad \left\{ \because P(r) = \frac{e^{-\lambda}\lambda^r}{r!} \right.$

$P(2) = \frac{2}{2} P(1) = .2706$

$P(3) = \frac{2}{3} P(2) = .1804$

$P(4) = \frac{1}{2} P(3) = .0902.$

Scanned by CamScanner

Q The frequency of accidents per shift in a factory is given in the following table

Accidents Per shift :	0	1	2	3	4
Frequency :	192	100	24	3	1

Calculate the mean number of accidents per shifts. find Corresponding Poisson distribution.

Solⁿ: Mean number of accidents per shift $= \frac{\Sigma x_i f_i}{\Sigma f_i}$

$$\lambda = \frac{100 + 2\times 24 + 3\times 3 + 4}{320} = 0.503$$

Theoretical frequency distribution will be as follows

X	$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Theoretical Frequency $N \times P(X)$
0	0.6047	193.5
1	0.3042	97.3
2	0.0765	24.5
3	0.0128	4.1
4	0.0016	0.5
		Total 319.9

Poisson distribution:

The Moment generating function about origin is

$$M_X(t) = E(e^{tX}) = \sum_r e^{tx_r} P(r) = \sum_{r=0}^{\infty} e^{tr} e^{-\lambda} \frac{\lambda^r}{r!}$$

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

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

$$\boxed{M_X(t) = e^{\lambda(e^t - 1)}}$$

Moments about origin:

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

$$\mu_1' = \text{mean} = \left[\frac{d}{dt} e^{\lambda(e^t-1)}\right]_{t=0}$$

$$= \left[\lambda e^t e^{\lambda(e^t-1)}\right]_{t=0}$$

$$\boxed{\mu_1' = \lambda} = \bar{x} = \text{mean}$$

$$\mu_2' = \left[\frac{d^2 M}{dt^2}\right]_{t=0} = \lambda\left[e^t e^{\lambda(e^t-1)} + \lambda e^{2t} e^{\lambda(e^t-1)}\right]_{t=0} = \lambda(1+\lambda)$$

$$\boxed{\mu_2' = \lambda^2 + \lambda}$$

$$\mu_3' = \left[\frac{d^3 M_X(t)}{dt^3}\right]_{t=0}$$

$$\boxed{\mu_3' = \lambda^3 + 3\lambda^2 + \lambda}$$

$$\mu_4' = \left[\frac{d^4 M_X(t)}{dt^4}\right]_{t=0}$$

$$\boxed{\mu_4' = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda}$$

Central Moments:

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

$$\mu_2 = \mu_2' - \mu_1'^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$$\boxed{\mu_2 = \lambda}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\boxed{\mu_3 = \lambda}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$\boxed{\mu_4 = 3\lambda^2 + \lambda}$$

Moment Generating function about $\bar{x}$ (mean):

$$M_X(t) \text{ about mean} = E[e^{t(X-\bar{x})}]$$
$$= E[e^{t(X-\lambda)}]$$
$$= e^{-\lambda t} E[e^{tX}]$$
$$= e^{-\lambda t} M_X(t) \text{ about origin}$$
$$= e^{-\lambda t} e^{\lambda(e^t - 1)} = e^{\lambda(e^t - 1 - t)}$$

$$\boxed{M_X(t) = e^{\lambda(e^t - 1 - t)}}$$

Moments about mean can be calculated by MGF about $\bar{x}$

$$\mu_1 = \left[\frac{d}{dt} M_X(t) \text{ (about mean)}\right]_{t=0} = 0$$

and so on.

Recurrence Relation for the central moments of Poisson Distribution

we have r^{th} moment about mean

$$\mu_r = E\{(X-\bar{x})^r\} = \sum_i P_i (x_i - \bar{x})^r$$
$$= \sum_{x=0}^{\infty} (x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{---} (1)$$

differentiate (1) w.r.to λ, we get,

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} (-r)(x-\lambda)^{r-1} \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!}\left(-e^{-\lambda}\lambda^x + x\lambda^{x-1}e^{-\lambda}\right)$$

$$= (-r)\sum_{x=0}^{\infty} (x-\lambda)^{r-1} \cdot \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!} \cdot e^{-\lambda}\lambda^x\left(-1 + \frac{x}{\lambda}\right)$$

$$= -r\sum_{x=0}^{\infty} (x-\lambda)^{r-1} P(x) + \frac{1}{\lambda}\sum_{x=0}^{\infty} (x-\lambda)^{r+1} P(x)$$

$$\frac{d\mu_r}{d\lambda} = -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

$$\boxed{\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}}$$

Karl Pearson's coefficient of Poisson distribution :

$$\boxed{\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda}}$$

$$\boxed{\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}}$$

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\lambda^2+\lambda}{\lambda^2} = 3+\frac{1}{\lambda}}$$

$$\boxed{\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}}$$

Distribution Function of Poisson Distribution :

$$F(x) = P(X \leq x) = \sum_{i=0}^{x} P(X=i)$$

or $F(x) = \sum_{r=0}^{x} P(X=r)$

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

Probability Generating Function of Poisson distribution :

$$G_X(z) = \sum_{r=0}^{\infty} \frac{e^{-\lambda}\lambda^r}{r!} z^r$$

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

$$\boxed{G_X(z) = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}}$$

Mode of Poisson Distribution :

Value of r for which $P(X=r)$ is maximum.

Now $P(r) > P(r-1)$ and $P(r) > P(r+1)$

Now $\frac{P(r)}{P(r-1)} > 1$

$$\Rightarrow \frac{\frac{e^{-\lambda}\lambda^r}{r!}}{\frac{e^{-\lambda}\lambda^{r-1}}{(r-1)!}} > 1 \quad \Rightarrow \frac{\lambda}{r} > 1 \text{ or } \lambda > r \text{ —— } (1)$$

and $\frac{P(r)}{P(r+1)} > 1 \Rightarrow \frac{(r+1)}{\lambda} > 1$ or $(r+1) > \lambda$ —— (2).

or $r > (\lambda - 1)$

From (1) and (2)

$$\boxed{\lambda-1 < r < \lambda}$$

Tutorial sheet

Q1 In a certain factory turning out razor blades, the probability of a blade to be defective is 0.01, The blades are sold in packet of 10. Use Poisson's distribution to find ~~probabilities~~ of a packet with

(i) No defective blade.
(ii) One blade (defective)
(iii) Two defective blades

Find the number of such packets in a consignment of 10,000 packets.

Sol^n we have

$n = 10, \quad p = 0.01$

$\lambda = np$

$\Rightarrow \lambda = 0.1$

(i) $P(X=0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-0.1} = 0.905$

(ii) $P(X=1) = \frac{e^{-\lambda}\lambda^1}{1!} = e^{-0.1}(0.1) = 0.0905$

(iii) $P(X=2) = \frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-0.1}(0.01)}{2} = 0.00452$

No. of packets with 0 defective blades $= 10{,}000 \times 0.905$
$= 9050$

No. of packets with 1 defective blade $= 10{,}000 \times 0.0905$
$= 905$

No. of packets with 2 defective blades $= 10{,}000 \times 0.00452$
$= 45$

Q.2 Records show that the probability is 0.00005 that a car will have a flat tyre while crossing a certain bridge. Use Poisson distribution to find probabilities that among 10,000 cars crossing this bridge,
(i) exactly two will have a flat tyre.
(ii) at most two will have a flat tyre.

Sol^n Let random variable X denote number of cars having flat tyres, which is a Poisson variate.

Here $n = 10{,}000$, $p = 0.00005$

Hence mean $= np = 0.5 = \lambda$

(i) $P(X=2) = \dfrac{e^{-\lambda}\lambda^2}{2!} = \dfrac{e^{-0.5}(0.5)^2}{2!} = \dfrac{(0.6065)(.25)}{2} = 0.0758$

(ii) $P(X \le 2) = P(0) + P(1) + P(2)$

$= e^{-0.5}\left[1 + 0.5 + \dfrac{(0.5)^2}{2}\right]$

$= (0.6065)(1.625) = 0.98556$

Q3 In a Poisson distribution if $3P(X=3) = 4P(X=4)$. Find $P(X=7)$

Solⁿ $3P(X=3) = 4P(X=4)$

$\dfrac{3e^{-\lambda}\lambda^3}{3!} = 4\dfrac{e^{-\lambda}\lambda^4}{4!}$

or $\lambda = 3$

$\therefore P(X=7) = \dfrac{e^{-3}3^7}{7!} = \dfrac{(0.04979)(2187)}{5040} = 0.0216$

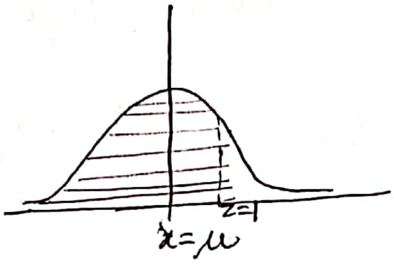
Q4 Evaluate the probabilities at $\mu = 10$ and $\sigma = 5$

(i) $P(X<15)$ (ii) $P(X \ge 15)$ (iii) $P(10 \le X \le 15)$

Solⁿ Here $\mu = 10$ and $\sigma = 5$

at $x = 15$

$z = \dfrac{x-\mu}{\sigma} = \dfrac{15-10}{5} = 1$



(i) $\Rightarrow P(X<15) = P(z<1)$

$= 0.5 + P(0<z<1)$

$= 0.5 + 0.3413$

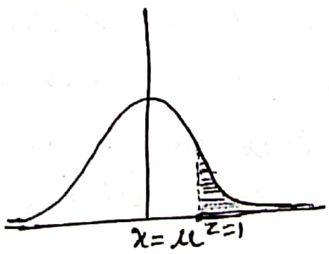
$= 0.8413$

(ii) $P(X \ge 15) = P(z \ge 1)$

$= 1 - P(z<1)$

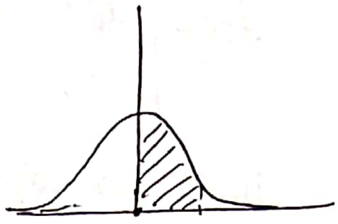
$= 1 - 0.8413$

$= 0.1587$



(iii) $P(10 < X \leq 15) = P(0 \leq Z \leq 1)$

$= 0.3413$



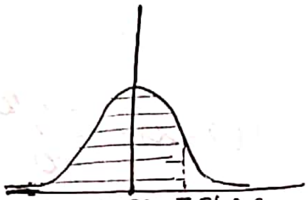
Q5 If X is normally distributed then find

(i) $P(Z \leq 1.26)$ (ii) $P(Z \geq 1.6)$ (iii) $P(0.2 \leq Z \leq 1.4)$

Soln (i) $P(Z \leq 1.26) = 0.5 + P(0 \leq Z \leq 1.26)$

$= 0.5 + 0.3962$

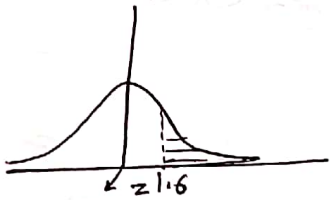
$= 0.8962$



(ii) $P(Z \geq 1.6) = 0.5 - P(0 < Z < 1.6)$

$= 0.5 - 0.4452$

$= 0.0548$



(iii) $P(0.2 \leq Z \leq 1.4) = P(0 \leq Z \leq 1.4) - P(0 \leq Z \leq 0.2)$

$= 0.4192 - 0.0793$

$= 0.3399$

Q.6 If X is uniformly distributed with mean 1 and variance $\frac{4}{3}$. Find $P(X < 0)$

Soln: $\because$ mean $= \frac{a+b}{2}$ and variance $= \frac{(b-a)^2}{12}$. [For uniform distribution]

or $1 = \frac{a+b}{2}$ and $\frac{4}{3} = \frac{(b-a)^2}{12}$

$\Rightarrow$ $a+b = 2$

& $(b-a) = 4$

on solving we have

$a = -1, \quad b = 3$ [we must have $a<b$]

Hence pdf of X is given by

4 distribution

$$f(x) = \begin{cases} \frac{1}{4}, & -1 < x < 3 \\ 0, & \text{otherwise} \end{cases} \quad \left[\because f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \right]$$

Hence $P(X<0) = \int_{-1}^{0} \frac{1}{4} dx = \frac{1}{4}$

Q7 If the mean of the Poisson distribution is 4, find $P(\lambda - 2\sigma < X < \lambda + 2\sigma)$

Sol^n. For a Poisson distribution variance $= \sigma^2 = \lambda$

Mean $= \mu = \lambda = 4$

$\Rightarrow \sigma = 2$

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-4}4^x}{x!}$$

$$P(\lambda-2\sigma < X < \lambda+2\sigma) = P(0 < X < 8)$$

$$= \sum_{x=1}^{7} P(X=x)$$

$$= \sum_{x=1}^{7} \frac{e^{-4}4^x}{x!}$$

$$= 0.9306$$

Q.8 Fit a Poisson distribution to the following data

Number of deaths (x) :	0	1	2	3	4
Frequency (f) :	122	60	15	2	1

Sol^n Mean $\frac{\Sigma f_i x_i}{\Sigma f_i} = 0.5$, For a Poisson distribution, mean $= \lambda = 0.5$

X	Theoretical Frequencies $N \times P(x) = 200 \times \frac{e^{-\lambda}\lambda^x}{x!}$
0	$121.31 \approx 121$
1	$60.65 \approx 61$
2	$15.16 \approx 15$
3	$2.53 \approx 3$
4	$0.32 \approx 0$

Q In a binomial distribution, the sum and product of the mean and variance are $\frac{25}{3}$ and $\frac{50}{3}$ respectively. Determine the distribution.

Sol^n For the binomial distribution,

$$np + npq = \frac{25}{3}$$

$$\text{or } np(1+q) = \frac{25}{3} \quad \text{--- (1)}$$

and $$np(npq) = \frac{50}{3}$$

or $$n^2p^2q = \frac{50}{3} \quad \text{--- (2)}$$

From (1) and (2),

$$\frac{n^2p^2(1+q)^2}{n^2p^2q} = \frac{625}{9} \times \frac{3}{50}$$

$$\frac{1+q^2+2q}{q} = \frac{25}{6}$$

or $$6q^2 - 13q + 6 = 0$$

or $$(2q-3)(3q-2) = 0$$

or $$q = \frac{3}{2} \text{ or } \frac{2}{3}$$

$\because$ q can not be greater than 1,

$\therefore$ $q = \frac{2}{3}$

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

From (1), $n = 15$

Hence, the binomial distribution is,

$$P(X=x) = {}^{15}C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{15-x}, \quad x = 0, 1, 2 \ldots 15$$

Q1. If the random variable X takes the values 1, 2, 3 and 4 such that $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4)$
Find the probability distribution and distribution function

Soln: Assume $P(X=4) = K$ then by definition of pmf

$\Sigma p_i = 1$

$\Rightarrow \left[\frac{5}{2} + \frac{5}{3} + 5 + 1\right] K = 1$

or $\frac{61}{6} K = 1$

or $K = \frac{6}{61}$

Now $P(X=1) = \frac{5}{2}K = \frac{5}{2} \cdot \frac{6}{61}$

$\boxed{P(X=1) = \frac{15}{61}}$

$P(X=2) = \frac{5}{3}K = \frac{5}{3} \times \frac{6}{61}$

$\boxed{P(X=2) = \frac{10}{61}}$

$\boxed{P(X=3) = 5K = 5 \times \frac{6}{61} = \frac{30}{61}}$

Hence required probability distribution is,

X :	1	2	3	4
$p(x_i)$:	$\frac{15}{61}$	$\frac{10}{61}$	$\frac{30}{61}$	$\frac{6}{61}$
$F(x)$:	$\frac{15}{61}$	$\frac{25}{61}$	$\frac{55}{61}$	1 (distribution function)

or $F(x) = \begin{cases} \frac{15}{61} ; & x \le 1 \\ \frac{25}{61} ; & x \le 2 \\ \frac{55}{61} ; & x \le 3 \\ 1 ; & x \le 4 \end{cases}$

~~or F(x) =~~ ~~0 ; -∞ < x < 1~~ ~~; 1 ≤ x < 2~~

Q2 Two cards are drawn without replacement from a well shuffled deck of 52 cards. Determine the probability distribution of the number of face cards.

Soln: Let X denotes the number of face cards. (i.e. Jack, Queen, King, Ace) obtained in a draw of 2 cards. Then $X = 0, 1, 2$,

A deck of 52 cards contains 16 face cards and 36 other cards.

$\therefore$ $P(X=0)$ = probability that no face card is obtained

$$= \frac{{}^{36}C_2 \; {}^{16}C_0}{{}^{52}C_2} = \frac{36}{52} \times \frac{35}{51} = \frac{105}{221}$$

$$P(X=1) = \frac{{}^{36}C_1 \; {}^{16}C_1}{{}^{52}C_2} = \frac{36 \times \cancel{35}}{52 \times 51} \times \frac{16 \times \cancel{15} \times 2}{1} = \frac{96}{221}$$

$$P(X=2) = \frac{{}^{36}C_0 \; {}^{16}C_2}{{}^{52}C_2} = \frac{16 \times 15}{52 \times 51} = \frac{20}{221}$$

(* Some authors consider 16 face cards like game Rummy. Some others consider 12 face cards)

Hence the required probability distribution is

X :	0	1	2
$P(X)$:	$\frac{105}{221}$	$\frac{96}{221}$	$\frac{20}{221}$

Q3 The probability distribution of a random variable X is given by

x_i :	0	1	2
p_i :	$3c^3$	$4c-10c^2$	$5c-1$

where $c>0$

Find (i) c, (ii) $P(X<2)$; (iii) $P(1<X\leq 2)$

Soln: $\because \Sigma p_i = 1$

$\Rightarrow 3c^3+4c-10c^2+5c-1=1$

$3c^3-10c^2+9c-2=0$

or $(c-2)(3c^2-4c+1)=0$

or $c = 1, 2, 1/3$.

But $\because 0\leq p_i\leq 1$

$\therefore$ the values $c=1$ and $c=2$ are not acceptable,

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

(ii) $P(X<2) = 1-P(X=2) = 1-(5c-1) = \frac{1}{3}$

(iii) $P(1<X\leq 2) = P(X=2) = 5c-1 = \frac{2}{3}$.

Q4 A random variable X has the following probability distribution.

$P(\frac{W}{A}) = \frac{P(A\cap W)}{P(A)}$

x :	0	1	2	3	4	5	6	7
$p(x)$:	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(i) Find k

(ii) Evaluate $P(X<6)$, $P(X\geq 6)$, $P(0<X<5)$.

(iii) Determine distribution function of X

(iv) If $P(X\leq c)>\frac{1}{2}$ find the minimum value of c.

(v) Find $P\left(\frac{1.5<X<4.5}{X>2}\right)$

Soln: (i) $\because \Sigma p_i = 1$

$\Rightarrow 10k^2+9k-1=0$

$k=-1, \frac{1}{10}$

T.13
R.V.

$k=-1$ is not possible as it makes $p(x)<0$ which is impossible, as above given is a probability distribution.

Hence $k=\frac{1}{10}$

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

$= 1-[P(X=6)+P(X=7)]$

$= 1-(9k^2+k) = 1-\frac{1}{10}-\frac{9}{100} = 1-\frac{19}{100} = \frac{81}{100}$

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

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

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

$= 8k = \frac{8}{10} = \frac{4}{5}$

(iii)

x_i :	0	1	2	3	4	5	6	7
$F(x)$:	0	$k=\frac{1}{10}$	$3k=\frac{3}{10}$	$5k=\frac{5}{10}$	$8k=\frac{8}{10}$	$8k+k^2 = \frac{81}{100}$	$8k+3k^2 = \frac{83}{100}$	1

(iv) $F(3) = P(X\leq 3) = 0.5$

$F(4) = P(X\leq 4) = 0.8 > \frac{1}{2}$

$F(5) = P(X\leq 5) = .81 > \frac{1}{2}$, and so on

Hence the minimum value of c for which $P(x\leq c) > \frac{1}{2}$ is 4

$\therefore c=4$

(v) $P\left(\frac{1.5<X<4.5}{X>2}\right) = \frac{P[(1.5<X<4.5)\cap(X>2)]}{P(X>2)}$

$= \frac{P(2<X<4.5)}{1-P(X\leq 2)} = \frac{P(3)+P(4)}{1-[P(0)+P(1)+P(2)]}$

$= \frac{\frac{2}{10}+\frac{3}{10}}{1-[0+\frac{1}{10}+\frac{2}{10}]} = \frac{\frac{5}{10}}{1-\frac{3}{10}} = \frac{5/10}{.7/10} = \frac{5}{7}$

Q5 Let X be a continuous random variable with p.d.f.

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

(i) Determine the constant a

(ii) Find $P(X \le 1.5)$

(iii) Determine the cdf and hence find $P(X \le 2.5)$

Soln (i) As $f(x)$ is given to be a pdf, hence

$$\int_0^\infty f(x)\,dx = 1$$

$$\Rightarrow \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^3 (-ax+3a)\,dx = 1$$

$$\Rightarrow \frac{a}{2} + a + \left(-\frac{a}{2}5\right) + 3a = 1$$

$$4a - 2a = 1$$

$$2a = 1 \quad \text{or} \quad \boxed{a = \frac{1}{2}}$$

(ii) $$P(X \le 1.5) = \int_0^{1.5} f(x)\,dx$$

$$= \int_0^1 ax\,dx + \int_1^{1.5} a\,dx = \frac{a}{2} + (0.5)a = \frac{1}{4} + \frac{0.5}{2}$$

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

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

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

(iii) For $x < 0$, $F(x) = 0$

For $0 \le x \le 1$; $F(x) = \int_0^x f(x)\,dx = \int_0^x ax\,dx = \frac{ax^2}{2} = \frac{x^2}{4}$

For $1 \le x \le 2$; $F(x) = \int_0^1 f(x)\,dx + \int_1^x f(x)\,dx$

$$= \int_0^1 ax\,dx + \int_1^x a\,dx$$

$$= \frac{a}{2} + a(x-1) = \frac{1}{4} + \frac{(x-1)}{2}$$

$$F(x) = \frac{x}{2} - \frac{1}{4}$$

for $2 \le x \le 3$; $F(x) = \int_0^1 f(x)\,dx + \int_1^2 f(x)\,dx + \int_2^x f(x)\,dx$

$$= \int_0^1 ax\,dx + \int_1^2 a\,dx + \int_2^x (-ax+3a)\,dx$$

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

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

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

for $x \ge 3$; $F(x) = \int_0^\infty f(x)\,dx = 1$

i.e. $$F(x) = \begin{cases} 0\,; & x < 0 \\ \frac{x^2}{4}\,; & 0 \le x \le 1 \\ \frac{x}{2} - \frac{1}{4}\,; & 1 \le x \le 2 \\ -\frac{5}{4} + \frac{3}{2}x - \frac{x^2}{4}\,; & 2 \le x \le 3 \\ 1\,; & x \ge 3 \end{cases}$$

and $$P(X \le 2.5) = F(2.5) = -\frac{5}{4} + \frac{3}{2}(2.5) - \frac{(2.5)^2}{4} = 1.25 + 3.75 - 1.5625$$

$$= 3.75 - 2.8125 = 0.9375$$

Q1 The Joint Probability mass function of (X,Y) is given by
$p(x,y) = K(2x+3y)$, $x = 0,1,2$; $y = 1,2,3$ find

(i) K
(ii) Marginal probability distribution of X
(iii) Marginal probability distribution of Y.
(iv) Conditional distribution of X given $Y=1$
(v) Conditional distribution of Y given $X=2$
~~(vi) the probability distribution of (X+Y)~~

Sol^n: The joint probability distribution of (X,Y) can be represented in tabular form as :-

X \ Y	1	2	3	Total
0	3K	6K	9K	18K
1	5K	8K	11K	24K
2	7K	10K	13K	30K
Total	15K	24K	33K	72K

(i) As above given is a pmf, hence

$$\sum_{i=0}^{2} \sum_{j=1}^{3} p(x_i, y_j) = 1$$

$$72K = 1$$

$$K = \frac{1}{72}$$

(ii) Marginal probability distribution of X is given by

$$p_{i}^{*} = p(x_i) = \sum_{j} p(x_i, y_j)$$

X	p_i^*
0	18/72
1	24/72
2	30/72

(iii) marginal probability distribution of Y

$$p_{j}^{*} = \sum_{i} p(x_i, y_j)$$

Y	p_j^*
1	15/72
2	24/72
3	33/72

(iv) Conditional distribution of X given Y=1 is

$$P\left(\frac{X=x_i}{Y=1}\right) = \frac{p(x_i,1)}{P(Y=1)} = \frac{p(x_i,1)}{p_{*j}} \quad \text{where } p_{*j} = \frac{15}{72} \text{ (at } Y=1) = 15K$$

X	$P\left(\frac{X=x_i}{Y=1}\right)$
0	$3K/15K = \frac{1}{5}$
1	$\frac{5K}{15K} = \frac{1}{3}$
2	$\frac{7K}{15K} = \frac{7}{15}$

(v) Conditional distribution of Y given X=2

$$P\left(\frac{Y=y_j}{X=2}\right) = \frac{P(x=2,y_j)}{P(X=2)} = \frac{p_{ij}}{p_{i*}} \quad \text{where } p_{i*} = P(X=2) = \sum_j p_{ij} = \frac{30}{72} = 30K$$

Y	$P\left(\frac{Y=y_j}{X=2}\right)$
1	$\frac{7K}{30K} = \frac{7}{30}$
2	$\frac{10K}{30K} = \frac{10}{30}$
3	$\frac{13K}{30K} = \frac{13}{30}$

Q The joint pdf of the random variable (X,Y) is given by

$$f(x,y) = kxy\, e^{-(x^2+y^2)}, \quad x>0,\ y>0$$

Find 'k' and prove also that X and Y are independent.

Soln: ∵ Given function is a pdf

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

or $$k\int_0^{\infty}\int_0^{\infty} xy\, e^{-x^2} e^{-y^2}\,dx\,dy = 1$$

or $$k\int_0^{\infty} xe^{-x^2}\left(\frac{e^{-y^2}}{-2}\right)_0^{\infty} dx = 1$$

$$\frac{k}{2}\int_0^{\infty} xe^{-x^2}\,dx = 1$$

or $$\frac{k}{4} = 1 \Rightarrow \boxed{k=4}$$

Marginal density of X

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$
$$= \int_0^{\infty} 4\,xy\, e^{-x^2} e^{-y^2}\,dy$$
$$= 4\,x e^{-x^2} \left(\frac{e^{-y^2}}{-2}\right)_0^{\infty}$$
$$f_X(x) = 2x e^{-x^2}, \quad x > 0$$

Marginal density of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$
$$= \int_0^{\infty} 4xy\, e^{-x^2} e^{-y^2}\,dx$$
$$= 4y e^{-y^2} \left(\frac{e^{-x^2}}{-2}\right)_0^{\infty}$$
$$f_Y(y) = 2y e^{-y^2}, \quad y > 0$$

Now $f_X(x)\,f_Y(y) = 4xy\, e^{-(x^2+y^2)}$

$= f(x,y), \quad x>0,\ y>0$

Hence X and Y are independent R.V.

Q3 Let X be a random variable with the following probability distribution.

X	:	-3	6	9
$p(X=x)$	:	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Find $E(X)$, $E(X^2)$, $E(2X+1)^2$

Soln $E(X) = \sum_i x_i p_i = -\frac{1}{2} + 3 + 3 = \frac{11}{2}$

$E(X^2) = \sum_i x_i^2 p_i = \frac{3}{2} + 18 + 27 = \frac{93}{2}$

$E(2X+1)^2 = E(4X^2+4X+1) = 4E(X^2) + 4E(X) + 1$

$= 4\left(\frac{93}{2}\right) + 4\left(\frac{11}{2}\right) + 1 = 186 + 22 + 1 = 209$

Q4 A bag contains 2 one rupee coin and 3, 50 paise coins. A person is allowed to draw two coins indiscriminately. Find the expected value of the draw.

Solⁿ: Let random variable X denote the amount drawn in rupees. The pmf of R.V. X is

X :	1	1.50	2
$P(x)$:	$\frac{{}^3C_2}{{}^5C_2} = \frac{3}{10}$	$\frac{{}^2C_1 \times {}^3C_1}{{}^5C_2} = \frac{6}{10}$	$\frac{{}^2C_2}{{}^5C_2} = \frac{1}{10}$

Hence $E(X) = 1 \times \frac{3}{10} + (1.50)\frac{6}{10} + 2 \times \frac{1}{10} = \frac{14}{10} = 1.40$ Rs.

Lecture - 1

Mathematical Expectation and Theoretical distributions

Random Variable: A random variable X is a function $X: S \rightarrow R$ that assigns a real number $X(s)$ to each $s \in S$ (Sample ~~sap~~ space), corresponding to a random experiment E.

eg① If we toss two coins together, we may consider the random variable X which is number of heads.

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

$$X(HH) = 2$$
$$X(HT) = 1$$
$$X(TH) = 1$$
$$X(TT) = 0$$

Note: Random variables are generally denoted by capital letters X, Y, Z etc.

eg② A single fair dice is rolled and the random variable X represents the number that turns up. Hence X can take values 1, 2, 3, 4, 5, 6.

eg.③ Two balls are drawn in succession without replacement from an urn containing 4 white and 3 green balls. X is the number of white balls, the values x of the random variable X are:

Sample space	x
WW	2
WG	1
GW	1
GG	0

We observe that a random variable is a variable whose value is determined by the outcome of a random experiment.

Note: Countably infinite :- If a set consists of points which can be arranged in to a simple sequence $s_1, s_2 \ldots$ otherwise it is known as uncountable.

eg. set of ~~whole~~ natural no. $= \{0, 1, 2, 3 \ldots\}$ Countably infinite
Integers $= \{\ldots, -2, -1, 0, 1, 2, 3 \ldots\}$ Countably infinite

set of Real no. $= \mathbb{R}$ uncountable

• **Discrete Random Variable:** A discrete random variable has either finite or countably infinite number of values.

eg. Let X be a random variable which denotes the number of heads in two successive tosses of a fair coin.
∴ Sample Space $= \{HH, HT, TH, TT\}$.

$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0$.

Hence X is a discrete random variable.

(**Note:** Sample space → finite or countably infinite then associated random variable will be discrete. and S is uncountable or continuous. then associated random variable ~~will be continuous.~~ may be discrete. eg. $S = \{s :$ height of individuals in a large group$\}$ and let random variable X denote height in inches rounded to nearest whole number than we can have $X(s) = 59$ inches. i.e. X is a discrete random variable.)

Note: Important discrete distributions are uniform distribution, Binomial distribution, Negative binomial dist., Hypergeometric dist., Poisson dist.

Continuous Random Variable: Which can take infinite number of values in an interval.

eg. temprature, current, voltage etc., the height of a group of individuals.

Note: Some random variable is neither a discrete nor a continuous known as mixed random vari...

Scanned by CamScanner

Discrete Probability distribution of a Random Variable :

Let X be a random variable with possible values x_i and associated probabilities $p(x_i)$; $i=1,2\dots,n$. Then the set with elements having the ordered pairs $(x_i, p(x_i))$ forms a probability distribution of a random variable X, where $p(x_i)$ has to satify the following conditions, :

(a) $p(x_i) \geq 0, \; \forall x_i$

(b) $\sum_{i=1}^{n} p(x_i) = p(x_1)+p(x_2)+\dots+p(x_n)=1$

is called the discrete probability distribution for X.

i.e.

x_i :	x_1	x_2	x_3	$\dots$	x_n
$p(x_i)$:	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\dots$	$p(x_n)$.

eg. When tossing a coin and denoting random variable X as the number of heads obtained, the probability distribution is :

$X=x$:	0	1
$P(x)$:	$\frac{1}{2}$	$\frac{1}{2}$.

{ sample space = (H, T).

Note: Here p_i is said to be probability mass function (pmf).

eg. ① Probability mass function

$$P(X=x) = \frac{x^2}{25} \;\forall\; x=1,2,3,4$$

x :	1	2	3	4
$P(X=x)$:	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{9}{25}$	$\frac{16}{25}$

$P(X=x) > 0 \;\forall\; x$, and $\sum_{x=1}^{4} P(X=x) = \frac{6}{5} > 1$

Probability distribution of a continuous Random Variable :

In case of continuous random variable, instead of finding the probability at a particular value of x we find the probability of x in a small interval.

We define the continuous probability distribution of x by $f(x)$ i.e. $P\left(x-\frac{dx}{2} < x < x+\frac{dx}{2}\right) = f(x)\,dx$

The continuous curve $y=f(x)$ is known as probability curve.

Probability Density function: The function $f(x)$ for a continuous random variable X is said to be probability density function (p.d.f.) provided it satisfies the following conditions.

1) $f(x) \geq 0 \ \forall;\ -\infty < x < \infty$

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

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

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

$\because P(X=a) = P(a \leq X \leq a) = \int_a^a f(x)\,dx = 0$

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

Soln: $f(x) > 0 \ \forall\ 0 \leq x \leq 1$

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

Mathematical Expectation: The expectation of a random variable X is defined as

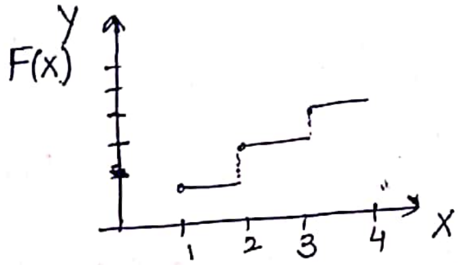
$$\bar{X} = E(X) = \begin{cases} \sum_i x_i p_i, & \text{if } X \text{ is discrete Random variable with pmf } p_i \\ \int_{-\infty}^{\infty} x f(x)\,dx, & \text{if } X \text{ is continuous RV with pdf } f(x) \end{cases}$$

Distribution Function (Discrete Random Variable):

~~Let X~~. The distribution function $F(x)$ of the discrete random variable X, is defined as

$$F(x) = P(X \leq x) = \sum_{i=1}^{n} p(x_i) \text{ where } x_1 \leq x, x_2 \leq x, x_3 \leq x \ldots, x_n \leq x$$

$F(x)$ is also known as cummulative distribution function i.e. cdf



Graph of distribution function

Properties of distribution function

(i) domain of distribution function is $(-\infty, \infty)$ and range is $[0,1]$

(ii) $F(x)$ treated as a step function.

(iii) If $x_1 \geq x_2$ then $F(x_1) \geq F(x_2)$

(iv) $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1)$
$= F(x_2) - F(x_1) = \sum_{i=1}^{i=2} P(X = x_i)$

(v) $F(-\infty) = 0$ and $F(\infty) = 1$

(vi) F is constant in the interval (x_k, x_{k+1}), it takes a jump of size $\{P(X = x_{k+1}) - P(X = x_k)\}$ at x_{k+1}

(vii) $F(X) = F(x_k) \ \forall \ x_k \leq X < x_{k+1}$
and $F(x_{k+1}) = F(x_k) + P(X = x_{k+1})$

Q In a supply of 10 similar T.V.s by a company. 4 are known to be defective. A college purchases 3 TVs from this company. Find the probability distribution for the number of defective TVs purchased and distribution function.

Soln: If the random variable X denote the number of defective TVs then X can take the values 0, 1, 2, 3, therefore

$p(0) = P(X=0) = \frac{{}^{4}C_0\,{}^{6}C_3}{{}^{10}C_3} = \frac{20}{120} = \frac{1}{6}$; $F(x=0) = \frac{1}{6}$

$p(1) = P(X=1) = \frac{{}^{4}C_1\,{}^{6}C_2}{{}^{10}C_3} = \frac{60}{120} = \frac{1}{2}$; $F(x=1) = \frac{1}{6}+\frac{1}{2} = \frac{2}{3}$

$p(2) = P(X=2) = \frac{{}^{4}C_2\,{}^{6}C_1}{{}^{10}C_3} = \frac{36}{120} = \frac{3}{10}$; $F(x=2) = \frac{1}{6}+\frac{1}{2}+\frac{3}{10} = \frac{29}{30}$

$p(3) = P(X=3) = \frac{{}^{4}C_3\,{}^{6}C_0}{{}^{10}C_3} = \frac{4}{120} = \frac{1}{30}$; $F(x=3) = \frac{1}{6}+\frac{1}{2}+\frac{3}{10}+\frac{1}{30} = 1.$

The probability distribution $p(x)$ of X and the distribution function $F(x)$ are given by.

X :	0	1	2	3
$p(x)$:	1/6	1/2	3/10	1/30
$F(x)$:	1/6	2/3	29/30	1

eg Q. Consider experiment of three tosses of a coin and consider the random variable X as the number of heads. Find Probability distribution and distribution function.

Soln: Sample space for this experiment.

$S = \{HHH, TTT, HHT, HTT, HTH, THT, THH, TTH\}$

X;	0	1	2	3
$p(x)$;	1/8	3/8	3/8	1/8

$$F(x) = \begin{cases} P(X=0) = 1/8 & ;\text{when } X=0 \\ P(X=0)+P(X=1) = 4/8 & ;\ X \le 1 \\ P(X=0)+P(X=1)+P(X=2) = 7/8 & ;\ X \le 2 \\ P(0)+P(X=1)+P(X=2)+P(X=3) = 1 & ;\ X \le 3 \end{cases}$$

Distribution Function (Continuous Random Variable) :

Let X be a continuous random variable having probability density function f(x), then $F_X(x)$ will be a continuous distribution function of X if

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

distribution function is also known as cummulative distribution function.

Properties of continuous Distribution function :

(i) $0 \le F_X(x) \le 1$; $-\infty < x < \infty$

(ii) $F_X(-\infty) = \lim_{x \to -\infty} F(x) = \int_{-\infty}^{-\infty} f(x)\,dx = 0$

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

(iii) $P(x_1 \le X \le x_2) = P(X \le x_2) - P(X \le x_1)$
$= F(x_2) - F(x_1)$.

Similarly.

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

(iv) $P(X = a) = \int_{a}^{a} f(x)\,dx = 0$

(v) $\frac{d}{dx} F_X(x) = f(x)$ or $F(x) = \int f(x)\,dx$

Q consider the function

$$f(x) = \begin{cases} c & ; a \le X \le b \\ 0 & ; \text{elsewhere} \end{cases}$$

(a) For what value of c, f(x) is a p.d.f.
(b) Find the distribution function of X

$\frac{\mathcal{L}_1}{R.V.}$

Sol^n (a) $f(x)$ will be a p.d.f. if

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

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

$$\Rightarrow \int_a^b c\,dx = 1 \Rightarrow c(b-a) = 1$$

$$\Rightarrow c = \frac{1}{(b-a)}.$$

(b) The distribution function

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

$$= \begin{cases} 0 & ; \ x < a \\ \frac{x-a}{(b-a)} & ; \ a \le x < b \\ 1 & ; \ x \ge b \end{cases}$$

$(\because$ for $x \ge b$ $F(x) = \int_0^{\infty} f(x)\,dx = 1)$

Q The distribution function of the random variable X is given by.

$$F_X(x) = \begin{cases} 0 & ; \ x < 2 \\ \alpha(x-2) & ; \ 2 \le x < 6 \\ 1 & ; \ x \ge 6 \end{cases}$$

(i) Find α (ii) $P(X > 4)$ (iii) $P(3 \le X \le 5)$

Sol^n (i) : $F_X(6) = 1$

$\alpha(6-2) = 1$

$\Rightarrow \alpha = \frac{1}{4}$

(ii) $P(X > 4) = 1 - P[X \le 4]$

$= 1 - F_X(4) = 1 - \frac{1}{4}(4-2) = \frac{1}{2}$

(iii) $P[3 \le X \le 5] = F_X(5) - F_X(3) = \frac{1}{4}[(5-2) - (3-2)] = \frac{1}{2}$

for $1 \leq x \leq 3$

Q Find pdf of a random variable X whose cdf is given by

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ x & , \quad 0 \leq x \leq 1 \\ 1 & , \quad x > 1 \end{cases}$$

Solⁿ

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

As $F(x)$ is not differentiable at $x=0$ and $x=1$ hence we can define $f(x)=0$ for $x=0$ and $x=1$.

Q (i) Is the function, defined as follows, a density function?

$$f(x) = \begin{cases} e^{-x} & , \quad x \geq 0 \\ 0 & , \quad x < 0 \end{cases}$$

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

(iii) Also find the cummulative probability function F(2).

Solⁿ: (i) Clearly, $f(x) \geq 0$ for every x in (1,2) and

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

Hence the function $f(x)$ satisfies the requirements for a density function.

(ii) Required Probability $= \int_1^2 f(x)\,dx = \int_1^2 e^{-x}\,dx = e^{-1} - e^{-2}$

$= 0.368 - 0.135 = 0.233$

(iii) $F(2) = \int_{-\infty}^{2} f(x)\,dx = \int_0^2 e^{-x}\,dx = 1 - e^{-2} = 1 - 0.135$

$= 0.865$

Scanned by CamScanner

Q Show that

$$F(x) = \begin{cases} 0 & ; \quad x \leq -a \\ \frac{1}{2}\left(\frac{x}{a}+1\right) & ; \quad -a \leq x \leq a \\ 1 & ; \quad x > a \end{cases}$$

is a distribution function.

Solⁿ: $F(x)$ will be distribution function if

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

(iii) $\frac{d}{dx}[F(x)] = f(x) \geq 0$ where $\int_{-\infty}^{\infty} f(x)dx = 1$, are satisfied i.e.

$f(x)$ must be p.d.f.

$\because F(x)=0$ when $x < -a$

and $F(x)=1$ when $x>a$

So conditions (i) and (ii) are satisfied.

Again $\frac{d}{dx}[F(x)] = f(x) = \begin{cases} 1/2a & \text{if, } -a \leq x \leq a \\ 0 & \text{if, otherwise} \end{cases}$

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

Thus condition (iii) is also satisfied. Hence $F(x)$ is distribution function.

Bivariate Random Variable : Let S be the sample space associated with the random experiment E. Then the function $f: S \to R^2$ where $f(s) = (X, Y)$, where $s \in S$ is said to be a Two dimensional random variable.

(i) If X and Y are discrete random variable Then (X, Y) is called discrete bivariate random Variable.

(ii) If X and Y are continuous random variable then (X, Y) is called a continuous bivariate random variable.

(iii) If one of X and Y is discrete and the other is continuous then (X, Y) is called a mixed bivariate random variable.

Discrete Bivariate Random Variable :

(1) **Joint Probability Mass Function :**

$$P_{XY}(x,y) = P_{XY}(X = x_i, Y = y_j) = P_{ij}, \text{ where}$$

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

(ii) $\sum_i \sum_j P_{ij} = 1$

(2) **Joint Distribution Function (Cumulative Distribution Function) (cdf) :**

$$F_{XY}(x,y) = P\{X \leq x, Y \leq y\}$$

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

Or $$F_{XY}(x,y) = \sum_{i=-\infty}^{x} \sum_{j=-\infty}^{y} P(X = i, Y = j)$$

(3) **Properties of Joint Distribution Function :**

(i) $0 \leq F_{XY}(x,y) \leq 1$

(ii) $F_{XY}(-\infty, -\infty) = 0$

(iii) $F_{XY}(\infty, \infty) = 1$

(iv) $F_{XY}(-\infty, Y) = F_{XY}(X, -\infty) = 0$

(V) $P(x_1 \leq X \leq x_2, Y \leq y) = F_{XY}(x_2, y) - F_{XY}(x_1, y)$

(VI) $P(X \leq x, y_1 \leq Y \leq y_2) = F_{XY}(x, y_2) - F_{XY}(x, y_1)$

(VII) $P(x_1 < X \leq x_2, y_1 \leq Y \leq y_2) = F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - \{F_{XY}(x_2, y_1) - F_{XY}(x_1, y_1)\}$

Continuous Bivariate R.V. :

(1) Joint Probability density function and joint distribution function:

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$

where $F_{XY}(x,y)$ is the joint distribution function defined as

$$F_{XY}(x,y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y)\, dx\, dy$$

and f_{XY} is joint probability density function if

(a) $f_{XY}(x,y) \geq 0$, $-\infty < x < \infty$, $-\infty < y < \infty$.

(b) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy = 1$

(Discrete)

Marginal Probability Distribution : When we are concerned with more than one random variable, the pmf/ probability distribution of a single variable is referred to as marginal pmf/ probability.

If we consider two-dimensional random variable (X,Y) then, the marginal probability function of X is defined as

$$P(X = x_i) = \sum_{j} p_{ij} = p_{i1} + p_{i2} + \dots + p_{in} + \dots = p_i^*$$

and the collection of pairs $\{x_i, p_i^*\}$ $i = 1,2 \dots, m \dots$ is called the marginal probability distribution of X.

Similarly $\{y_j, p_j^*\}$ $j = 1,2 \dots, n, \dots$ is called the marginal probability distribution of Y.

(Discrete)

Conditional Probability Distribution : Consider two dimensional discrete R.V. (X,Y) The conditional probability function of X, given $Y = y_j$ is given by.

$$P\left\{\frac{X = x_i}{Y = y_j}\right\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_j^*}$$

and $\left\{x_i, \frac{p_{ij}}{p_j^*}\right\}$, $i = 1,2 \dots m, \dots$ is called the conditional probability distribution of X, given $Y = y_j$

Similarly, the conditional probability function of Y, given $X = x_i$ is given by

$$P\left\{\frac{Y=y_j}{X=x_i}\right\} = \frac{P\{X=x_i, Y=y_j\}}{P\{X=x_i\}} = \frac{p_{ij}}{p_{i^*}}$$

and $\left\{ y_j, \frac{p_{ij}}{p_{i^*}} \right\}$ is called the conditional probability distribution of Y given $X = x_i$.
where $j = 1, 2 \ldots, n, \ldots$

<u>Independent Random variables (discrete)</u> : Let (X, Y) be two dimensional random variable such that

$$P\left\{\frac{X=x_i}{Y=y_j}\right\} = P\{X = x_i\}$$

i.e. $\frac{p_{ij}}{p_{*j}} = p_{i^*}$

i.e. $p_{ij} = p_{i^*} \cdot p_{*j}$ $\forall$ i and j, then X and Y are said to be independent random variables.

<u>Q</u> The joint distribution function of a random variable (X, Y) is given by. $F_{XY}(x, y) = \begin{cases} (1-e^{-ax})(1-e^{-by}) & ; x, y \geq 0, a, b > 0 \\ 0 & ; \text{otherwise} \end{cases}$

Find (i) Marginal distribution function of X and Y
(ii) $P(X \leq 2, Y \leq 2)$ and $P(X \leq 1)$.
Also show that X and Y are independent

<u>Sol^n</u> By the definition of marginal distribution function

(i) $F_X(x) = F_{XY}(x, \infty)$
and $F_Y(y) = F_{XY}(\infty, y)$

$$\therefore F_X(x) = F_{XY}(x, \infty) = \begin{cases} (1-e^{-ax}) & ; x \geq 0. \\ 0 & ; \text{otherwise}. \end{cases}$$

$$F_Y(y) = F_{XY}(\infty, y) = \begin{cases} (1-e^{-by}) & ; y \geq 0. \\ 0 & ; \text{otherwise} \end{cases}$$

(ii) $\because F_{XY}(x, y) = P(X \leq x, Y \leq y)$

$\therefore P(X \leq 2, Y \leq 2) = F_{XY}(2, 2) = (1-e^{-2a})(1-e^{-2b})$

also $P(X \le 1) = F_X(1) = (1 - e^{-9})$

Again two random variable X and Y are independent if $F_{XY}(x,y) = F_X(x)\,F_Y(y)$ using the two results of (i) we get the required result.

~~Q The joint probability mass function of (X, Y) is given by.~~

~~$P_{XY}(x_i, y_j) = \begin{cases} \lambda x_i^2 y_j \;;\; i=1,2 \;;\; j=1,2,3. \\ 0 \;;\; \text{otherwise} \end{cases}$~~

~~Find (i) λ~~
~~(ii) the marginal probability mass function of X and Y.~~

Soln. Q The joint probability distribution of random variable (X, Y) is given by $P_{XY}(x_i, y_j) = \frac{1}{27}(x_i + 2y_j)$ where x and y can assume only the integer values 0, 1, 2.

(i) Find the marginal distribution of X and Y.
(ii) Are X and Y independent ?
(iii) Find the conditional distribution of Y for X=2.

Soln As (X, Y) is a bivariate random variable. Then in tabular form $P_{XY} = \frac{1}{27}(x_i + 2y_j)$ can be represent as follows.

X \ Y	y_1 0	y_2 1	y_3 2	Total
x_1 0	0	2/27	4/27	6/27
x_2 1	1/27	3/27	5/27	9/27
x_3 2	2/27	4/27	6/27	12/27
Total	3/27	9/27	15/27	1

(i) The marginal probability mass function of X

$$P_X(x_i) = \sum P_{XY}(x_i, y_j)$$
$$= \sum_{j=0}^{y_2} P_{XY}(x_i, y_j)$$

X=i	$P_X(x_i)$
0	6/27
1	9/27
2	12/27
Total	1

Again marginal distribution mass function for Y is

$P_Y(y_j) = \sum_i P_{XY}(x_i, y_j)$

$Y=j$	$P_Y(x_j)$
0	3/27
1	9/27
2	15/27
Total	1

(ii) Two random variable X and Y are called independent if, $P_{XY}(x_i, y_j) = P_X(x_i)\, P_Y(y_j)$

But from the table it is clear

$P_X(X=0) = \frac{6}{27}$ and $P_Y(Y=0) = \frac{3}{27}$

$\therefore P_X(x=0)\, P_Y(y=0) = \frac{6}{27} \times \frac{3}{27} = \frac{2}{81} \neq 0 = P_{XY}(x=0, y=0)$

Hence two variables are not independent.

(iii) Conditional distribution of Y given $X=2$ is

$P_Y\left\{\frac{Y=y_j}{X=x_i}\right\} = \frac{P_{XY}(x_i, y_j)}{P_X(x_i)}$

where $P_X(x=2) = \sum_j P(x=2, y_j)$

$= \frac{2}{27} + \frac{4}{27} + \frac{6}{27} = \frac{12}{27}$

$Y=j$	$P\left(\frac{y_j}{x_i}\right)$
0	$\frac{2/27}{12/27} = \frac{1}{6}$
1	$\frac{4/27}{12/27} = \frac{1}{3}$
2	$\frac{6/27}{12/27} = \frac{1}{2}$

Q Two balls are selected at random from a box containing two red, three white and four blue balls. Let (X,Y) be a bivariable random variable where X and Y denotes the number of red and white balls chosen.

(i) Find joint probability mass function of (X,Y).
(ii) Find marginal probability mass function of X and Y.
(iii) Conditional distribution of X given $Y=1$.
(IV) Are X and Y independent R.V.

Solⁿ: According to problem X and Y denotes the number of red and white balls choosen. So X and Y taken values 0, 1, 2 subject to the condition $X+Y \geq 0$.

Total number of balls = 9

So the number of ways of drawing two balls from the bag are

$$^9C_2 = \frac{9.8}{2.1} = 36$$

∴ The various probabilities are

$$P_{XY}(0,0) = \frac{^3C_0 \times {^2C_0} \times {^4C_2}}{36} = \frac{1}{6} \quad ; \quad P_{XY}(1,2) = 0$$

$$P_{XY}(0,1) = \frac{^2C_0 \times {^3C_1} \times {^4C_1}}{36} = \frac{1}{3} \quad ; \quad P_{XY}(2,1) = 0$$

$$P_{XY}(0,2) = \frac{^2C_0 \times {^3C_2} \times {^4C_0}}{36} = \frac{1}{12} \quad ; \quad P_{XY}(2,2) = 0$$

$$P_{XY}(1,0) = \frac{^2C_1 \times {^3C_0} \times {^4C_1}}{36} = \frac{2}{9}$$

$$P_{XY}(1,1) = \frac{^2C_1 \times {^3C_1} \times {^4C_0}}{36} = \frac{1}{6}$$

$$P_{XY}(2,0) = \frac{^2C_2 \times {^3C_0} \times {^4C_0}}{36} = \frac{1}{36}$$

X \ Y	0	1	2	Total
0	1/6	1/3	1/12	7/12
1	2/9	1/6	0	7/18
2	1/36	0	0	1/36
Total	15/36	1/2	1/12	1

(ii) Now the marginal distribution of x.

$$P_X(x_i) = \sum_{j=0}^{2} P_{XY}(x_i, y_j)$$

$$P_Y(y_j) = \sum_{i=0}^{2} P_{XY}(x_i, y_j)$$

$X=i$	$P_X(x_i)$
0	7/12
1	7/18
2	1/36

$Y=j$	$P_X(y_j)$
0	15/36
1	1/2
2	1/12

(iii) Conditional distribution of X given $Y = 1$

$$P\left\{\frac{X = x_i}{Y = y_j}\right\} = \frac{P_{XY}(x_i, y_j)}{P_Y(y_j)}$$

$X = x_i$	$P_X\left(\frac{x_i}{y_j}\right)$
0	$\frac{1/3}{1/2} = \frac{2}{3}$
1	$\frac{1/6}{1/2} = \frac{1}{3}$
2	$\frac{0}{1/2} = 0$

(IV) Two variables X and Y are called independent if

$$P_{XY}(x_i, y_j) = P_X(x_i)\, P_Y(y_j)$$

From the table $P_X(0) = \frac{7}{12}$ and $P_Y(0) = \frac{15}{36}$

$$\therefore\ P_X(0)\, P_Y(0) = \frac{7}{12} \times \frac{15}{16} = \frac{35}{144} \neq \frac{1}{6} = P_{XY}(0,0)$$

Hence two variables are not independent.

(conti.)

Marginal density: Let (X, Y) be a two dimensional continuous random variable. Then

marginal density of X is $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$

marginal density of Y is $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$.

Note: $P(a \leq X \leq b) = P(a \leq X \leq b, -\infty < Y < \infty)$

$$= \int_{-\infty}^{\infty} \int_a^b f(x,y)\, dx\, dy = \int_a^b \left[\int_{-\infty}^{\infty} f(x,y)\, dy\right] dx$$

$$= \int_a^b f_X(x)\, dx$$

$$P(c \leq Y \leq d) = \int_c^d f_Y(y)\, dy$$

(conti.)

Conditional Density: Let (X, Y) be a two dimensional continuous random variable. Then the conditional density of X given Y denoted by $f\left(\frac{x}{y}\right)$ is given by.

$$f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f_Y(y)}$$

Similarly, the conditional density of Y given X, is given by

$$f\left(\frac{y}{x}\right) = \frac{f(x,y)}{f_X(x)}$$

Independent Continuous Random Variables:

Let (X,Y) be a two-dimensional continuous random variable then X and Y are said to be independent random variable if
$$f(x,y) = f_X(x)\, f_Y(y).$$

Q. Assume that the lifetime X and the brightness Y of a light bulb are being modeled as continuous random variables with joint pdf given by
$$f(x,y) = \lambda_1 \lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}, \quad 0 < x < \infty, \; 0 < y < \infty.$$
Find the joint distribution function.

Soln: · The joint distribution function is given by.
$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy$$
$$= \int_0^y \left[\int_0^x \lambda_1\lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}\,dx\right] dy.$$
$$= \lambda_1\lambda_2 \int_0^y e^{-\lambda_2 y}\left(\frac{e^{-\lambda_1 x}}{-\lambda_1}\right)_0^x dy$$
$$= \lambda_2 (1 - e^{-\lambda_1 x}) \cdot \left(\frac{e^{-\lambda_2 y}}{-\lambda_2}\right)_0^y.$$
$$F(x,y) = (1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2 y}) \quad ; \quad 0 < x < \infty, \; 0 < y < \infty.$$

Q The joint probability density function of a bivariate random variable (X,Y) is given by
$$f_{XY}(x,y) = \begin{cases} \lambda(x+y) & ; \quad 0 < x < 3, \; 0 < y < 3. \\ 0 & ; \quad \text{otherwise} \end{cases}$$
where λ is a constant

(i) Find the value of λ

(ii) Find the marginal probability density function of X and Y.

(iii) Are X and Y independent ?

Sol^n By the definition of joint probability density function

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

$\Rightarrow \lambda \int_0^3 \int_0^3 (x+y)\,dx\,dy = 1$

$\lambda \int_0^3 \left(\frac{x^2}{2} + xy\right)_0^3 dy = 1$

$\Rightarrow \lambda \cdot \int_0^3 \left(\frac{9}{2} + 3y\right) dy = 1$

$\Rightarrow \lambda\left(\frac{9}{2}y + \frac{3}{2}y^2\right)_0^3 = 1$

$\Rightarrow \lambda\left[\frac{27}{2} + \frac{27}{2}\right] = 1$

$\Rightarrow \boxed{\lambda = \frac{1}{27}}$

(ii) Again by the definition of marginal probability density function of X given Y=y.

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

$= \frac{1}{27}\int_0^3 (x+y)\,dy$

$= \frac{1}{27}\left[xy + \frac{y^2}{2}\right]_0^3 = \frac{1}{27}\left[3x + \frac{9}{2}\right]$

$\Rightarrow f_X(x) = \begin{cases} \frac{1}{54}(6x+9) & ; \ 0<x<3 \\ 0 & ; \text{ otherwise} \end{cases}$

Similarly the marginal probability density function of Y given X = x

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

$= \frac{1}{27}\int_0^3 (x+y)\,dx$

$f_Y(y) = \begin{cases} \frac{1}{54}(6y+9) & ; \ 0<y<3 \\ 0 & ; \text{ otherwise} \end{cases}$

(iii) It is clear from the given function and from the Case (ii) $f_{XY}(x,y) \neq f_X(x)\, f_Y(y)$

Hence X and Y are not independent random variable.

Q. The joint pdf of the random variable (X, Y) is given by.

$$f(x,y) = kxy\, e^{-(x^2+y^2)}, \quad x>0,\; y>0.$$

Find 'k' and prove also that X and Y are independent.

Solⁿ. By the definition of joint probability density function

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = k\int_{0}^{\infty}\int_{0}^{\infty} xy\, e^{-(x^2+y^2)}\,dx\,dy = 1$$

$$\Rightarrow k\int_0^{\infty} y e^{-y^2}\left\{\int_0^{\infty} x e^{-x^2}dx\right\}dy = 1$$

$$\Rightarrow k\int_0^{\infty} y e^{-y^2}\left(\frac{-e^{-x^2}}{2}\right)_0^{\infty} dy = 1$$

$$\Rightarrow \frac{k}{2}\int_0^{\infty} y e^{-y^2}\,dy = 1$$

$$\Rightarrow \frac{k}{4}\left(-e^{-y^2}\right)_0^{\infty} = 1$$

$$\Rightarrow \boxed{k=4}$$

Marginal density of $X = f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$$= \int_0^{\infty} 4xy\, e^{-(x^2+y^2)}\,dy$$

$$= 4x e^{-x^2}\left[\frac{-e^{-y^2}}{2}\right]_0^{\infty}$$

$$= 2x e^{-x^2}, \quad x>0.$$

Marginal density of $Y = f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

$$= \int_0^{\infty} 4xy\, e^{-(x^2+y^2)}\,dx$$

$$= 2y e^{-y^2}, \quad y>0$$

Now $f_X(x)\, f_Y(y) = 4xy\, e^{-(x^2+y^2)} = f(x,y), \quad x>0, y>0.$

Hence X and Y are independent random variables.

2.2.11.
R.V

Q. Given the joint probability density $f(x,y) = \begin{cases} \frac{2}{3}(x+2y) ; 0<x<1, 0<y<1 \\ 0 \quad ; \text{elsewhere} \end{cases}$

Find (i) Marginal density of X and Y
(ii) Conditional density of X given Y = y.
and use it to evaluate $P\left\{\frac{X \le 1/2}{Y = 1/2}\right\}$.

Sol: (i) Marginal density of X

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

$$= \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}\left(xy + \frac{2y^2}{2}\right)_0^1$$

$$= \frac{2}{3}(x+1) \quad ; \quad 0<x<1$$

similarly.
Marginal density of Y

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

$$= \int_0^1 \frac{2}{3}(x+2y)\,dx = \frac{2}{3}\left(\frac{x^2}{2} + 2xy\right)_0^1$$

$$f_Y(y) = \frac{2}{3}\left(\frac{1}{2} + 2y\right) = \frac{1}{3}(1+4y) \; ; \; 0<y<1$$

(ii) Conditional density of X given Y = y is

$$f\left(\frac{X}{Y}\right) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{2}{3}(x+2y)}{\frac{1}{3}(1+4y)} = \frac{2(x+2y)}{(1+4y)}, \quad 0<x<1$$

Hence, $P\left[\frac{X \le 1/2}{Y = \frac{1}{2}}\right] = \int_0^{1/2} f\left(\frac{x}{y=\frac{1}{2}}\right) dx$

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

$$= \frac{1}{3}\left[\frac{9}{4} - 1\right] = \frac{1}{3} \times \frac{5}{4} = \frac{5}{12}.$$

Exponential Distribution:

A random variable X is said to have an exponential distribution if its p.d.f. is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \ x \geq 0 \\ 0 & ; \ \text{elsewhere} \end{cases}$$

where λ is a parameter and $\lambda > 0$,

Clearly $\int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_0^{\infty} e^{-\lambda x}\,dx = 1$

Distribution Function of exponential distribution:

$$F(x) = P(X \leq x) = \int_0^x f(x)\,dx$$

$$= \int_0^x \lambda e^{-\lambda x}\,dx = (-e^{-\lambda x})_0^x$$

$$\boxed{F(x) = (1 - e^{-\lambda x})}, \quad x \geq 0$$

Moments, Moment Generating Function, Mean and Variance:

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_0^{\infty} e^{-(\lambda - t)x}\,dx$$

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

$$\boxed{M_X(t) = E(e^{tX}) = \frac{\lambda}{(\lambda - t)}, \text{ where } \lambda > t}$$

Also the moments about origin are given as:

$$\mu_r' = E(X^r) = \int_0^{\infty} x^r \lambda e^{-\lambda x}\,dx$$

let $\lambda x = y$

$dx = \frac{dy}{\lambda}$

$$= \frac{1}{\lambda^r} \int_0^{\infty} y^r e^{-y}\,dy$$

$$= \frac{1}{\lambda^r} \int_0^{\infty} e^{-y} y^{(r+1)-1}\,dy$$

$$\boxed{\mu_r' = E(X^r) = \frac{1}{\lambda^r} \Gamma(r+1) = \frac{r!}{\lambda^r}}$$

Hence $\boxed{\text{Mean } \mu_1' = \frac{1}{\lambda}}$

$$\boxed{\mu_2' = \frac{2}{\lambda^2}}$$

$$\sigma^2 = \text{Variance} = \mu_2' - \mu_1'^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Standard deviation $= \sigma = \frac{1}{\lambda}$

Memoryless Property of Exponential distribution:

If X is exponentially distributed, then

$$P\left(\frac{X > x_1+t}{X>t}\right) = P(X>x_1) \quad \forall\ x_1, t>0$$

$\because$ $P(X \le x) = (1-e^{-\lambda x})$

or $P(X>x) = e^{-\lambda x}$

Now $$P\left(\frac{X>x_1+t}{X>t}\right) = \frac{P(X>x_1+t \cap X>t)}{P(X>t)}$$

$$= \frac{P\{X>(x_1+t)\}}{P\{X>t\}} = \frac{e^{-\lambda(x_1+t)}}{e^{-\lambda t}}$$

$$= e^{-\lambda x_1}.$$

$$P\left(\frac{X>x_1+t}{X>t}\right) = P(X>x_1)$$

The converse of above result is also true. Hence, if $P\left(\frac{X>x_1+t}{X>t}\right) = P(X>x_1)$ then X follows exponential distribution.

Q The time (in hours) required to repair a machine is exponentially distributed with parameter $\lambda = \frac{1}{2}$.

(i) what is the probability that the repair time exceeds 2 hrs?

(ii) what is the conditional probability that a repair takes at least 10 hours given that its duration exceeds 9 hours?

Soln: Here X represents time (in hours) required to repair a machine, then its pdf is given as

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2} & ; x \ge 0 \\ 0 & ; \text{otherwise} \end{cases}$$

(i) $P(X>2\text{ hrs}) = \int_2^\infty f(x)dx = \frac{1}{2}\int_2^\infty e^{-x/2}dx = \frac{1}{e} = 0.3679$

(ii) $P(X \ge 10 / X>9) = P(X>1) = \int_1^\infty f(x)dx = \frac{1}{2}\int_1^\infty e^{-x/2}dx$

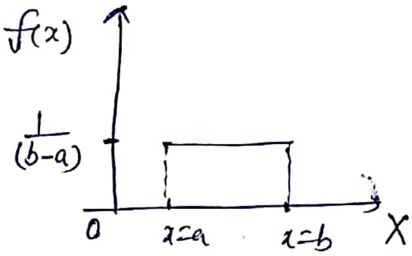
$= 0.6065$

Rectangular or Uniform Distribution :

A continuous random variable X is said to follow a continuous ~~dtt~~ uniform distribution over an interval (a, b), if its p.d.f is given by.

$$f(x) = \begin{cases} \frac{1}{(b-a)} & ; a < x < b \\ 0 & ; \text{elsewhere} \end{cases}$$

Here X is known as uniform variate with parameters a and b.



(Curve of Rectangular distribution)

Distribution Function of Rectangular Distribution :

$$F(x) = P(X \leq x) = \int_a^x f(x)\,dx$$

$$= \int_a^x \frac{1}{(b-a)}\,dx = \frac{x-a}{b-a}$$

$$F(x) = \frac{x-a}{(b-a)}, \quad \text{for } a \leq x \leq b$$

Moment Generating Function, Moments, Mean and Variance :

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_a^b \frac{e^{tx}}{(b-a)}\,dx$$

$$\boxed{M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}} \quad \underline{MGF}$$

Moments about origin

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\,dx = \frac{1}{b-a}\int_a^b x^r\,dx$$

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

$$\boxed{\therefore \mu_1' = \text{Mean} = \frac{a+b}{2}}$$

$\frac{2.}{U.D.}$

$$\mu_2' = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\therefore \text{Variance} = \sigma^2 = \mu_2' - \mu_1'^2 = \mu_2$$

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

$$\text{Standard Deviation} = \sigma = \frac{b-a}{2\sqrt{3}}$$

$z = \ z=0.2 \quad z=1.4$

Q.6 If X is uniformly distributed with mean 1 and variance $\frac{4}{3}$. Find $P(X<0)$

Soln:

$\because$ mean $= \frac{a+b}{2}$ and variance $= \frac{(b-a)^2}{12}$. [For uniform distribution]

or $1 = \frac{a+b}{2}$ and $\frac{4}{3} = \frac{(b-a)^2}{12}$.

$\Rightarrow a+b=2$

& $(b-a) = 4$

on solving, we have

$a=-1, \quad b=3$ [we must have $a<b$]

Hence pdf of X is given by

$$f(x) = \begin{cases} \frac{1}{4}, & -1 < x < 3 \\ 0, & \text{otherwise} \end{cases} \qquad \left[\because f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0; & \text{otherwise} \end{cases} \right]$$

4

distribution

Hence $P(X < 0) = \int_{-1}^{0} \frac{1}{4}\, dx = \frac{1}{4}$

Poisson distribution is 4 kind

Probability Distribution

Some discrete probability distributions are

a) Binomial Distribution

b) Poisson Distribution

P.T.O

Binomial Distribution

A random variable X in any experiment is said to follow Binomial distribution if its pmf is

$$P(X=r) = P(r) = {}^nC_r p^r q^{n-r} \quad r = 0, 1, 2 \ldots n$$
$$= (p+q)^n$$

where experiment is repeated n times. Each trial is independent and has only two outcomes success (p) or failure (q). $(q = 1-p)$

Mean and Variance of binomial Distribution

$$\text{Mean } \mu = E(X) = \sum_{r=0}^{n} x_r p(X=r) = \sum_{r=0}^{n} x_r p(r) \quad = \mu_1'$$

$$= \sum_{r=0}^{n} r\, {}^nC_r p^r q^{n-r}$$

$$= 0 + 1\, {}^nC_1 p q^{n-1} + 2\, {}^nC_2 p^2 q^{n-2} + 3\, {}^nC_3 p^3 q^{n-3} + \ldots + np^n]$$

$$= npq^{n-1} + \frac{2n(n-1)}{2} p^2 q^{n-2} + \frac{n(n-1)(n-2)}{\lfloor 2} p^3 q^{n-3} + \ldots + np^n$$

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

$$= np \sum_{r=0}^{n-1} {}^{n-1}C_r p^r q^{(n-1)-r} = np(p+q)^{n-1}$$

$$= np \qquad (p+q=1) \qquad \text{—} (1)$$

To find Variance we calculate μ_2'

$$\mu_2' = E(X^2) = \sum_{r=0}^{n} r^2\, {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} [r^2 - r + r]\, {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} [r(r-1) + r]\, {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} r(r-1)\, {}^nC_r\, p^r q^{n-r} + \sum_{r=0}^{n} r\, {}^nC_r\, p^r q^{n-r}$$

$$= \cancel{\sum \{n(n-1)} = 2(2-1)\, {}^nC_2\, p^2 q^{n-2} + 3(3-1)\, {}^nC_3\, p^3 q^{n-3} + \dots + n(n-1)\, {}^nC_n\, p^n q^0 + \sum_{r=0}^{n} r\, {}^nC_r\, p^r q^{n-r}$$

$$\Rightarrow 2\,\frac{n(n-1)}{2}\, p^2 q^{n-2} + 3\cdot 2\, \frac{n(n-1)(n-2)}{\lfloor 3}\, p^3 q^{n-3} + \dots$$

$$+\; n(n-1)\, p^n q^0 + \sum_{r=0}^{n} r\,p(X=r) \quad \text{(using ①)}$$

$$\Rightarrow n(n-1)p^2\left[q^{n-2} + (n-2)\,p\,q^{n-3} + \frac{(n-2)(n-3)}{\lfloor 2}\, p^2 q^{n-4} + \dots + p^{n-2}\right] + np$$

$$= n(n-1)p^2 \sum_{r=0}^{n-2} {}^{n-2}C_r\, p^r q^{n-2} + np$$

$$= n(n-1)p^2\, [q+p]^{n-2} + np \qquad (p+q=1)$$

$$= n(n-1)\,p^2 \cdot (1)^{n-2} + np$$

$$= np\,[(n-1)p + 1] = np\,(np+1-p) = np\,(np+q)$$

Variance $= E(X^2) - [E(X)]^2 = \mu_2' - (\mu_1')^2$

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

Variance $= npq$

Standard deviation $= \sqrt{Var} = \sqrt{npq}$

Moments and Moment Generating Function $M_X(t)$

The moment generating function about the origin is

$$M_X(t) = E[e^{tx}] = \sum_{r=0}^{n} e^{tr}\, {}^nC_r\, p^r q^{n-r} = \sum_{r=0}^{n} e^{xt} p(r)$$

$$= \sum_{r=0}^{n} e^{xt}\, {}^nC_r\, p^r q^{n-r}$$

$$= \sum_{r=0}^{n} {}^nC_r (pe^t)^r q^{n-r} \qquad = \sum_{r=0}^{n} {}^nC_r (pe^t)^r q^{n-r}$$

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

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

Now moments about origin

$$\mu_1' = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[npe^t(q+pe^t)^{n-1}\right]_{t=0}$$

$$= np(q+p) = np$$

$$\mu_2' = \left[\frac{d^2}{dt^2} M_X(t)\right] = \frac{d}{dt}\left[\frac{d}{dt} M_X(t)\right]_{t=0}$$

$$= \frac{d}{dt}\left[np\, e^t (q+pe^t)^{n-1}\right]_{t=0}$$

$$= np\left[e^t(q+pe^t)^{n-1} + (n-1)e^{2t} p (q+pe^t)^{n-2}\right]_{t=0}$$

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

$$= np\left[1 + p(n-1)\right] \qquad p+q=1$$

$$= np[1+np-p] = np[np+q] = n^2p^2 + npq$$

$$\mu_3' = \left[\frac{d^3}{dt^3} M_X(t)\right] = np\frac{d}{dt}\left[e^t(q+pe^t)^{n-1} + (n-1)e^{2t}p(q+pe^t)^{n-2}\right]_{t=0}$$

$$= np\left[e^t(q+pe^t)^{n-1} + e^t(n-1)pe^t(q+pe^t)^{n-2} + 2(n-1)p\, e^{2t}(q+pe^t)^{n-2} + (n-1)e^{3t}p(n-2)p(q+pe^t)^{n-3}\right]_{t=0}$$

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

$p+q=1$

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

$= np[1+3(n-1)p + (n-1)(n-2)p^2]$

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

Similarly $\mu_4' = \left[\frac{d^4}{dt^4} M_x(t)\right]$

$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$

Central Moments of Binomial distribution

Here we use the interrelation between μ_r and μ_r'

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = \cancel{n^2p^2 - np^2} =$

$n^2p^2 + npq - n^2p^2 = npq$

$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

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

$= np[1 + 3(n-1)p + (n-1)(n-2)p^2 - 3(n^2p^2 + npq) + 2n^2p^2]$

$= np[1 + 3np - 3p + (n^2 - 3n + 2)p^2 - n^2p^2 - 3npq]$

$= np[1 - 3p + 2p^2 + 3np - 3np^2 - 3npq]$

$= np[1 - 3p + 2p^2 + 3np(1-p) - 3npq]$

$= np[1 - 3p + 2p^2 + 3npq - 3npq]$

$= np[1 - 3p + 2p^2] = np(2p-1)(p-1) = np(1-p)(1-2p)$

$= npq(1-2p) = npq(1-p-p) = npq(q-p)$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

$$-4[np + 3n(n-1)p^2 + n(n-1)(n-2)p^3]np + 6(n^2p^2 + npq)(n^2p^2)$$
$$-3n^4p^4$$

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

Karl Pearson's β and γ coefficients for Binomial distribution

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{[npq(q-p)]^2}{n^3p^3q^3} = \frac{npq(q-p)^2}{npq} = \frac{(1-2p)^2}{npq}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{1-2p}{\sqrt{npq}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{npq[1+3(n-2)pq]}{n^2p^2q^2} = \frac{1+3(n-2)pq}{npq}$$

$$= \frac{1-6pq+3npq}{npq} = 3 + \frac{1-6pq}{npq}$$

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}$$

Probability Generating Function of Binomial Distribution

$$G_X(z) = \sum_{i=0}^{n} p_i z^i \qquad \text{as } i=r$$

$$= \sum_{r=0}^{n} {}^nC_r p^r q^{n-r} z^r = \sum_{r=0}^{n} {}^nC_r (pz)^r q^{n-r}$$

$$= (q+pz)^n$$

Recurrence Relation for the Central Moments

By definition $\qquad \mu = \text{mean} = np$

$$\mu_r = E[(X-\mu)^r] = \sum_{x=0}^{n} (x-np)^r P(x)$$

$$\mu_r = \sum_{x=0}^{n} (x-np)^r \, {}^nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} (x-np)^r \, {}^nC_x p^x (1-p)^{n-x}$$

diff. w.r.to 'p' we get

$$\frac{d\mu_r}{dp} = \sum_{x=0}^{n} {}^nC_x \left[-nr\,(x-np)^{r-1} p^x (1-p)^{n-x} \right.$$

$$\left. + (x-np)^{r} \left\{ x p^{x-1} (1-p)^{n-x} - p^x (n-x)(1-p)^{n-x-1} \right\} \right]$$

$$= \sum_{x=0}^{n} {}^nC_x (-nr)(x-np)^{r-1} p^x (1-p)^{n-x} + \sum_{x=0}^{n} (x-np)^{r} \, {}^nC_x$$

$$p^x q^{n-x} \left\{ \frac{x}{p} - \frac{n-x}{q} \right\}$$

$$= -nr \sum_{x=0}^{n} (x-np)^{r-1} P(x) + \sum_{x=0}^{n} (x-np)^r P(x) \left[\frac{xq - np + px}{pq} \right]$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq} \sum_{x=0}^{n} (x-np)^r P(x) \left[x(p+q) - np \right]$$

$$\because p+q=1$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq} \sum_{x=0}^{n} (x-np)^r P(x)\,(x-np)$$

$$= (-nr)\mu_{r-1} + \frac{1}{pq} \sum_{x=0}^{n} (x-np)^{r+1} P(x)$$

$$\frac{d\mu_r}{dp} = (-nr)\mu_{r-1} + \frac{1}{pq}\mu_{r+1}$$

$$\mu_{r+1} = pq\left[nr\,\mu_{r-1} + \frac{d\mu_r}{dp}\right]$$

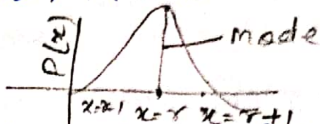
which is known as Recurrence relation for the central moments of Binomial distribution.

By putting $r = 1, 2, 3 \ldots$ we can find μ_2, μ_3 and μ_4 and so on.

Mode of Binomial distribution

Mode is the value of x for which $P(x)$ is maximum.

Let $X = r$ be the modal value.



$$\Rightarrow P(X=r) > P(X=r-1) \;\&\; P(X=r) > P(X=r+1)$$

Now
$$\frac{P(X=r)}{P(X=r+1)} = \frac{{}^nC_r\, p^r q^{n-r}}{{}^nC_{r+1}\, p^{r+1} q^{n-(r+1)}} = \frac{(r+1)\,q}{(n-r)\,p}$$

Since $P(X=r) > P(X=r+1) \Rightarrow \dfrac{P(X=r)}{P(X=r+1)} > 1$

$$\Rightarrow \frac{(r+1)\,q}{(n-r)\,p} > 1 \Rightarrow np - pr < qr + q$$

$$\Rightarrow np - q < qr + pr$$

$$np - q < (p+q)r$$

$$np - q < r$$

$$np - (1-p) < r \Rightarrow np - 1 + p < r \Rightarrow (n+1)p - 1 < r \quad \text{— (1)}$$

$$\frac{P(X=r)}{P(X=r-1)} = \frac{{}^nC_r\, p^r q^{n-r}}{{}^nC_{r-1}\, p^{r-1} q^{n-r+1}} = \frac{(n-r+1)\,p}{rq}$$

Since $P(X=r) > P(X=r-1) \Rightarrow \dfrac{P(X=r)}{P(X=r-1)} > 1$

$$(n-r+1)\,p > qr \Rightarrow (n+1)p > qr + pr$$

$$\Rightarrow np + p > (p+q)r \Rightarrow (n+1)p > r \quad \text{— (2)}$$

from ① & ②

~~np-q~~ $[(n+1)p-1] < r < (n+1)p$

Fitting of Binomial distribution

We have $\dfrac{P(r+1)}{P(r)} = \dfrac{{}^nC_{r+1}\, p^{r+1} q^{n-(r+1)}}{{}^nC_r\, p^r q^{n-r}} = \dfrac{\lfloor n}{\lfloor r+1 \lfloor n-(r+1)} \dfrac{\lfloor n-r \lfloor r}{\lfloor n} \dfrac{p}{q}$

$= \dfrac{(n-r)p}{(r+1)q}$

$\Rightarrow P(r+1) = \dfrac{(n-r)p}{(r+1)q} P(r)$

Fitting a binomial distribution means to find the theoretical frequencies for a given frequency distribution.

Mean of Binomial distribution $(np) = \dfrac{\sum_i x_i f_i}{N}$

Q.1 If during a war one out of 9 ships could not arrive safely. Find the probability that exactly 3 out of a convey of 6 would arrive safely.

Sol^n Let p (success)

q (failure)

$q = 1/9 \quad p = \dfrac{8}{9}$

probability of 3 arrive safely out of 6

${}^6C_3 p^3 q^3 = \dfrac{\lfloor 6}{\lfloor 3 \lfloor 3}\left(\dfrac{8}{9}\right)^3\left(\dfrac{1}{9}\right)^3 \qquad = \dfrac{10240}{96}$

Q.2 If 10% of pens manufactured by the company are defective, find the probability that a box of 12 pens contain

(i) Exactly two defective pens

(ii) Atleast two defective pens

(iii) No defective pen

Sol^n Let X denote no of defective pens

Here $n=12$ $p = \frac{10}{100} = 0.1$, $q = 1-p = 1-0.1 = 0.9$

$P(X=r) = {}^nC_r p^r q^{n-r}$

(i) $P(X=2) = {}^{12}C_2 (0.1)^2 (0.9)^{10} = 0.2301$

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

$= 1 - [{}^{12}C_0 (0.1)^0 (0.9)^{12} + {}^{12}C_1 (0.1)^1 (0.9)^{11}]$

$= 1 - [0.2824 + 0.3766]$

$= 1 - 0.659$

$= 0.341$

(iii) $P(X=0) = {}^{12}C_0 (0.1)^0 (0.9)^{12} = 0.2824$

Q.3 An irregular six faced dice is thrown and the probability that it gives five even numbers in 10 throws is twice the probability that it gives four even numbers in 10 throws. How many times in 10,000 sets of 10 throws each, would you expect to get no even number.

Sol^n Let X = no. of times an even no. is obtained

Let p = get an even no. , $n = 10$.

$P(X=r) = {}^nC_r p^r q^{n-r}$

Given $P(X=5) = 2P(X=4)$

${}^{10}C_5 p^5 q^5 = 2\, {}^{10}C_4 p^4 q^6$

$252p = (210)2q \Rightarrow 3p = 5q$

$\Rightarrow 3p = 5(1-p)$

$\Rightarrow 8p = 5 \Rightarrow p = 5/8,\ q = 3/8$

$P(X=r) = {}^{10}C_r \left(\frac{5}{8}\right)^r \left(\frac{3}{8}\right)^{10-r}$

$P(X=0) = \left(\frac{3}{8}\right)^{10} = 0.00005$

Q.4 Probability that a man aged 60 would be alive till 70 yrs of age is 0.65. Find the probability that atleast 7 out of 10 such men would be alive ~~70~~ till 70 years of age.

Sol^n X = no. of men aged 60 and would be alive till 70 yrs.

$n = 10, \; p = 0.65, \quad q = 1-p = 0.35$

$P(X = r) = {}^nC_r \, p^r q^{n-r}$

$P(X \geq 7) = P(7) + P(8) + P(9) + P(10)$

$= {}^{10}C_7 (0.65)^7 (0.35)^3 + {}^{10}C_8 (0.65)^8 (0.35)^2 + {}^{10}C_9 (0.65)^9 (0.35)$

$+ {}^{10}C_{10} (0.65)^{10}$

$= 120(0.00210) + 45(0.00390) + 10(0.00725) + 0.01346$

$= 0.252 + 0.1755 + 0.0725 + 0.01346 = 0.513$

Q.5 The following data gives the no. of seeds germinating out of 10 on damp filter paper for 80 sets of seeds. Fit a Binomial distribution to this data.

No. of seeds (x) :	0	1	2	3	4	5	6 and above
No. of sets (f) :	6	20	28	12	8	6	0

Sol^n Here $n = 10, \; \Sigma f_i = 80$

Mean $= \dfrac{\Sigma f_i x_i}{f_i} = \dfrac{174}{80} = 2.175 = np$ (mean)

$n = 10$ & $p = \dfrac{2.175}{10} = 0.2175 \quad q = 1-p = 0.7825$

Hence the binomial distribution to be approximated for this data $= N(p+q)^n$

$= 80\,(0.7825 + 0.2175)^{10}$

r	$P(r) = {}^nC_r p^r q^{n-r}$	$f(r) = NP(r) = 80P(r)$
0	$P(0) = q^{10} = (0.7825)^{10} = 0.08607$	$\cong 6.9$
1	$P(1) = {}^{10}C_1 (0.2175)(0.7825)^9 = 0.2392$	$\cong 19.1$
2	$P(2) = {}^{10}C_2 (0.2175)^2 (0.7825)^8 = 0.2992$	$\cong 23.9$
3	$P(3) = {}^{10}C_3 (0.2175)^3 (0.7825)^7 = 0.2218$	$\cong 17.7$
4	$P(4) = {}^{10}C_4 (0.2175)^4 (0.7825)^6 = 0.1079$	$\cong 8.6$
5	$P(5) = {}^{10}C_5 (0.2175)^5 (0.7825)^5 = 0.0359$	$\cong 2.9$
6	$P(6) = {}^{10}C_6 (0.2175)^6 (0.7825)^4 = 0.0083$	$\cong 0.8$
7	$P(7) = {}^{10}C_7 (0.2175)^7 (0.7825)^3 = 0.0013$	$\cong 0.1$
8 and above	negligible	$\cong 0$

8.6 Out of 800 families with 4 children each, how many family would be expected to have (i) 2 Boys and 2 girls (ii) atleast 1 boy (iii) at most 2 girls & (iv) children of both sex. Assume equal probabilities for boys and girls.

Sol^n Let X = number of girls.

$n = 4$. $N = 800$, $p = q = 1/2$ $\quad P(X=r) = {}^nC_r p^r q^{n-r}$

(i) $P(2\text{ boys } \& \ 2 \text{ girls}) = P(X=2) = {}^4C_2 p^2 q^2$

$$= {}^4C_2 \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^2 = 6 \times \left(\tfrac{1}{2}\right)^4 = \frac{3}{8}$$

(ii) No. of families having 2 boys and 2 girls $= 800 \times \frac{3}{8} = 300$

(ii) Let X = no. of boys.

$X = 0, 1, 2, 3$

$$P(X \geq 1) = 1 - P(X<1) = 1 - P(X=0)$$

$$1 - {}^4C_0 p^0 q^4 = 1 - \left(\tfrac{1}{2}\right)^4 = \frac{15}{16}$$

Total no. of families having atleast one boy $= \frac{15}{16} \times 800 = 750$

(iii) At most 2 girls $= P(X \leq 2)$

$$= P(X=0) + P(X=1) + P(X=2)$$
$$= {}^4C_0\left(\frac{1}{2}\right)^4 + {}^4C_1\left(\frac{1}{2}\right)^4 + {}^4C_2\left(\frac{1}{2}\right)^4$$
$$= \left(\frac{1}{2}\right)^4 [4 + 1 + 6] = \frac{11}{16}$$

No. of families having at most 2 girls $= 800 \times \frac{11}{16} = 550$

(iv) P(Children of both sexes) $= 1 - P($ all children are of same sex$)$

$$= 1 - [P\{\text{all are boys}\} + P\{\text{all are girls}\}]$$
$$= 1 - \{P(X=0) + P(X=4)\} = 1 - \{{}^4C_0 p^0 q^4 + {}^4C_4 p^4 q^0\}$$
$$= 1 - \left[\left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4\right] = 1 - \frac{2}{16} = 1 - \frac{1}{8} = \frac{7}{8}$$

No. of families having children of both sexes $= 800 \times \frac{7}{8} = 700$

Q 7 Find the parameters of the binomial distribution whose mean is 10 and variance 6.

Sol^n Let $X \sim B(n,p)$

Mean $= np = 10$ Variance $= npq = 6$

$$q = \frac{6}{10} = \frac{3}{5}, \quad p = 1 - q = 1 - \frac{3}{5} = \frac{2}{5}$$

$$np = 10 \Rightarrow n\frac{2}{5} = 10 \Rightarrow n = 25$$

$n = 25$, $p = 0.4$, $q = 0.6$

Q.8 In how many throws of a dice the probability of throwing 6 atleast once is just greater than 0.5.

Sol^n Let X = number of times 6 is obtained

$p = $(getting 6) $= \frac{1}{6} = p \Rightarrow q = \frac{5}{6}$

$P(X=r) = {}^nC_r p^r q^{n-r}$ $\quad r = 0, 1, 2 \ldots n$

Given $P(X \geq 1) > 0.5$

$1 - P(X=0) > 0.5 \Rightarrow P(X=0) < 0.5$

${}^nC_0 p^0 q^n < 0.5$

$\left(\frac{5}{6}\right)^n < 0.5 \Rightarrow n \log\left(\frac{5}{6}\right) < \log 0.5$

$n(-0.0792) < (-0.30103)$

$n(0.0792) > (0.30103)$

$n > \frac{0.30103}{0.0792} = 3.8$

$n = 4, 5, 6 \ldots$

Hence minimum no. of required throws $= 4$.

Q.8 8 coins are tossed simulltaneously 256 times. Number of heads observed at each throw are recorded and the results are given below. Find the expected frequency and a fit a binomial distribution. What are the theoretical values of the mean and standard deviation. Also calculate mean and stan. deviation of the observed frequencies

No. of Heads x	0	1	2	3	4	5	6	7	8	
No. of times f	2	6	30	52	67	56	32	10	1	$\Sigma 256$

Sol^n Observed Mean $= \mu_{obs} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1040}{256} = 4.0625$

$E(X^2) = \frac{\Sigma f_i x_i^2}{\Sigma f_i} = \frac{4772}{256} = 18.6406$

Observed variance $\sigma^2_{obs} = E(X^2) - [E(X)]^2$

$= 18.6406 - 16.5039 = 2.1367$

observed stand. deviation $= \sigma = 1.4617$

Scanned with CamScanner

Here $n = 8$ (no. of coins) $p = 1/2$ (head) $q = 1/2$

Theoretical mean $= np = 8 \times \frac{1}{2} = 4$

Theoretical stand. deviation $= \sqrt{npq} = \sqrt{8 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{2}$

$= 1.414$

Expected frequencies

$n = 8$ $np = 4.0625 \cong \Rightarrow p = 0.5078$ then $q = 1 - p$

$= 0.4922$

Expected frequencies are given by expansion of

$= N(p+q)^n$

$256\,(0.4922 + 0.5078)^8$

Q. 9 For a special security in a certain protected area, it was decided to put three lighting bulbs on each pole. If each bulb has a probability p of burning out in the first 100 hours of service, calculate the probability that atleast one of them is still good after 100 ~~years~~ hours. If $p = 0.3$, how many bulbs will be needed on each pole to ensure 99% safety so that at least one is good after 100 hours.

Sol^n: X = the no. of bulbs that do not burn out in first 100 hours.

Hence $P(\text{success}) = 1 - p$, $P(\text{failure}) = p$

$n = 3$

Required probability $= P(X \geq 1) = 1 - P(X = 0)$

$1 - {}^3C_0 (1-p)^0 p^3 = 1 - p^3$

$= 1 - (0.3)^3 = 0.973$

Now if $p = 0.3$, Let no. of bulbs on each pole be n to ensure 99% saftey so that atleast one is good after 100 hours.

$$P(X \geq 1) = 1 - P(X=0) = 1 - {}^nC_0 (1-p)^0 p^n$$
$$= 0.99$$

$$1 - (0.3)^n = 0.99$$

$$\Rightarrow 1 - 0.99 = (0.3)^n \Rightarrow n \log 0.3 = \log 0.01$$

$$\Rightarrow n = \frac{2}{0.5229} = 3.8 \simeq 4.$$

Q 10. The sum of mean and variance of a Binomial distribution is 15 and the sum of their squares is 117. Determine the distribution.

Sol^n Let n & p be the parameters of distribution

$$\text{Mean} = np \quad , \quad \text{variance} = npq$$

$$np + npq = \cancel{117}\ 15 \Rightarrow \text{on squaring}$$
$$np(1+q) = 15 \qquad n^2p^2(1+q)^2 = 225 \quad -①$$

and $n^2p^2 + n^2p^2q^2 = 117 \Rightarrow n^2p^2(1+q^2) = 117 \quad -②$

from ① & ②

$$\frac{(1+q)^2}{1+q^2} = \frac{225}{117} \Rightarrow \frac{q^2+2q+1}{1+q^2} = \frac{225}{117}$$

$$\frac{1+2q}{1+q^2} = \frac{225}{117} \Rightarrow \frac{2q}{1+q^2} = \frac{12}{13}$$

$$\frac{1+q^2}{2q} = \frac{13}{12} \Rightarrow \frac{1+q^2+2q}{1+q^2-2q} = \frac{13+12}{13-12} \quad \text{(by C \& D Rule)}$$

$$\frac{(1+q)^2}{(1-q)^2} = \frac{25}{1} \Rightarrow \frac{1+q}{1-q} = 5 \Rightarrow 6q = 4$$
$$q = 2/3$$

$$p = 1 - q = \frac{1}{3}$$

$np + npq = 15 \Rightarrow \frac{5n}{9} = 15$ (putting $p = \frac{1}{3}, q = \frac{2}{3}$)

$n = 27$

Hence the required distribution is

$P(X=r) = {}^{27}C_r \left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{27-r}$ $\quad r = 0, 1, 2 \ldots$

Q. 11 A Binomial variate satisfies the condition $9P(X=4) = P(X=2)$. If $n=6$ find p, $\bar{x}$ and σ.

Sol^n $\quad 9P(X=4) = P(X=2)$

$9\, {}^6C_4\, p^4 (1-p)^2 = {}^6C_2\, p^2 (1-p)^4$

$9p^2 = (1-p)^2 \Rightarrow 9p^2 = 1 + p^2 - 2p$

$8p^2 + 2p - 1 = 0$

$(4p-1)(2p+1) = 0 \Rightarrow p = 1/4, p = -1/2$

$p = -1/2$ is not possible so $p = 1/4$

Mean $= np = 6 \times \frac{1}{4} = \frac{3}{2}$

$\sigma = \sqrt{npq} = \sqrt{6 \cdot \frac{1}{4} \cdot \frac{3}{4}} = \sqrt{1.125} = 1.0607$

Lecture - 3

Poisson's Distribution: Poisson distribution is a limiting case of the binomial distribution under the following conditions.

1.) n, the number of trials is ~~iden~~ indefinitely large, i.e. $n \to \infty$
2.) p, the probability of success for each trial is indefinitely small, i.e. $p \to 0$
3.) $np = \lambda$ (say) is finite positive real number.

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

The probability of r success in a series of n independent trials is

$$P(r) = {}^nC_r\, p^r q^{n-r}$$

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

$$= \frac{n(n-1)\cdots(n-r+1)}{r!}\left(\frac{\lambda}{n}\right)^r \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left\{1-\frac{(r-1)}{n}\right\}\lambda^r}{r!}\,\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^r}$$

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(r) = \frac{\lambda^r e^{-\lambda}}{r!}, \quad r = 0, 1, \cdots$$

$$\left\{\because \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}\right\}.$$

This limiting form of Binomial distribution with above probability is called Poisson's distribution.

Note 1.) λ is known as the parameter of the distribution.

2.) $e = 2.7183$

3.) $\sum_{r=0}^{\infty} P(X=r) = \sum_{r=0}^{\infty} \frac{e^{-\lambda}\lambda^r}{r!} = e^{-\lambda}\left[1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right] = e^{-\lambda} e^{\lambda} = 1$

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non negative values and its probability mass function is given by

$$P(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x=0,1,2\ldots, \lambda>0 \\ 0, & \text{otherwise} \end{cases}$$

Note: This distribution is used to describe the behaviour of rare events such as the number of accidents on road, number of printing mistakes in a book etc.

Q. Suppose on an average 1 house in 1,000 in a certain district has a fire during a year. If there are 2,000 houses in that district, what is the probability that exactly 5 houses will have a fire during the year?

Solⁿ: $n = 2000$, $p = \frac{1}{1000}$

$\lambda = np = 2000 \times \frac{1}{1000} = 2$.

Required probability that exactly 5 houses will have a fire during the year $= P(5)$

$$= \frac{e^{-\lambda}\lambda^5}{5!}$$

$$= \frac{e^{-2}2^5}{5!}$$

$$= \frac{.135 \times 32}{120}$$

$$= .036$$

Mean and variance of the Poisson distribution:

for the Poisson distribution

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

$$E(X) = \text{Mean} = \mu = \sum_{r=0}^{\infty} r\,P(r)$$

$$= \sum_{r=0}^{\infty} r \frac{\lambda^r e^{-\lambda}}{r!}$$

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

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

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

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

$$= \lambda$$

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

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

$$= \sum_{r=0}^{\infty} r^2 P(X=r) - \lambda^2$$

$$= \sum_{r=0}^{\infty} r^2 \frac{\lambda^r e^{-\lambda}}{r!} - \lambda^2$$

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

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

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

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

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

$$= \lambda e^{-\lambda} \cdot e^{\lambda}(1+\lambda) - \lambda^2$$

$$= \lambda$$

Hence, standard deviation $\sigma = \sqrt{\text{var}(X)} = \sqrt{\lambda}$

Fitting a Poisson Distribution: When a Poisson distribution is to be fitted to observe data, the following procedure is adopted.

1.) Compute the mean $\bar{X}$ and take it equal to the mean of the fitted (Poisson) distribution.

$$\bar{X} = \lambda$$

2.) Obtain the probabilities $P(X=r) = \frac{e^{-\lambda}\lambda^r}{r!}, \ r = 0, 1, 2 \ldots$

3.) The expected or theoretical frequencies according to Poisson distribution can be calculated as

$$f(r) = N \cdot P(X=r)$$

where N is the total observed frequency.

Q. Data was collected over a period of 10 years, showing number of deaths from horse kicks in each of the 200 army corps. The distribution of deaths was as follows.

No. of deaths :	0	1	2	3	4	Total
Frequency :	109	65	22	3	1	$200 = N = \Sigma f_i$

Fit a Poisson distribution to the data and calculation the theoretical frequencies.

Sol^n.

x	f	fx
0	109	0
1	65	65
2	22	44
3	3	9
4	1	4
	$\Sigma f = 200$	$\Sigma f x = 122$

~~$\bar{x} = \Sigma f x$~~ $\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{122}{200} = 0.61 = \lambda$.

$\lambda \cdot 3 \cdot 5$

Math. Expec. and Theo. Distri.

X	$P(X=x) = \frac{e^{-\lambda} \lambda^x}{x!}$	Frequency $NP(X=x)$
0	$e^{-.61} \frac{(.61)^0}{0!} = .5432$	$200 \times .5432 = 108.64 \approx 109$
1	$e^{-.61} \frac{(.61)^1}{1!} = .3313$	$200 \times .3313 = 66.27 \approx 66$
2	$e^{-.61} \frac{(.61)^2}{2!} = .101$	$200 \times .101 = 20.21 \approx 20$
3	$e^{-.61} \frac{(.61)^3}{3!} = .021$	$200 \times .021 = 4.11 \approx 4$
4	$e^{-.61} \frac{(.61)^4}{4!} = .003$	$200 \times .003 = .63 \approx 1.$

Recurrence formula for the Poisson Distribution:

$$\because \quad P(r) = \frac{e^{-\lambda} \lambda^r}{r!}$$

$$P(r+1) = \frac{e^{-\lambda} \lambda^{r+1}}{(r+1)!}$$

$$\Rightarrow \quad P(r+1) = \frac{\lambda}{(r+1)} \cdot P(r).$$

Q If the variance of the Poisson distribution is 2, find the probabilities for $r = 1, 2, 3, 4$ from the recurrence relation of the Poisson distribution.

Soln: Here $\lambda = 2$

$\because \; P(r+1) = \frac{\lambda}{(r+1)} P(r) = \frac{2}{(r+1)} P(r)$. which is the recurrence relation

$P(1) = 2 \cdot P(0) = 2 \cdot e^{-2} = 2 \times .1353 = .2706$ $\quad \left\{ \because P(r) = \frac{e^{-\lambda} \lambda^r}{r!} \right.$

$P(2) = \frac{2}{2} P(1) = .2706$

$P(3) = \frac{2}{3} P(2) = .1804$

$P(4) = \frac{1}{2} P(3) = .0902.$

Q The frequency of accidents per shift in a factory is given in the following table

Accidents Per shift :	0	1	2	3	4
Frequency :	192	100	24	3	1

Calculate the mean number of accidents per shifts. find Corresponding Poisson distribution.

Solⁿ: Mean number of accidents per shift $= \frac{\sum x_i f_i}{\sum f_i}$

$$\lambda = \frac{100 + 2\times 24 + 3\times 3 + 4}{320} = 0.503$$

Theoretical frequency distribution will be as follows

X	$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$	Theoretical frequency $N \times P(X)$
0	0.6047	193.5
1	0.3042	97.3
2	0.0765	24.5
3	0.0128	4.1
4	0.0016	0.5
		Total 319.9

Poisson distribution:

The Moment generating function about origin is

$$M_X(t) = E(e^{tX}) = \sum_r e^{tx_r} P(r) = \sum_{r=0}^{\infty} e^{tr} e^{-\lambda} \frac{\lambda^r}{r!}.$$

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

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

$$\boxed{M_X(t) = e^{\lambda(e^t-1)}}$$

Moments about origin:

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

$$\mu_1' = \text{mean} = \left[\frac{d}{dt} e^{\lambda(e^t-1)}\right]_{t=0}$$

$$= \left[\lambda e^t e^{\lambda(e^t-1)}\right]_{t=0}$$

$$\boxed{\mu_1' = \lambda} = \bar{x} = \text{mean}$$

$$\mu_2' = \left[\frac{d^2 M}{dt^2}\right]_{t=0} = \lambda\left[e^t e^{\lambda(e^t-1)} + \lambda e^{2t} e^{\lambda(e^t-1)}\right]_{t=0} = \lambda(1+\lambda)$$

$$\boxed{\mu_2' = \lambda^2+\lambda}$$

$$\mu_3' = \left[\frac{d^3 M_X(t)}{dt^3}\right]_{t=0}$$

$$\boxed{\mu_3' = \lambda^3 + 3\lambda^2 + \lambda}$$

$$\mu_4' = \left[\frac{d^4 M_X(t)}{dt^4}\right]_{t=0}$$

$$\boxed{\mu_4' = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda}$$

Central Moments:

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

$$\mu_2 = \mu_2' - \mu_1'^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$$\boxed{\mu_2 = \lambda}$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$\boxed{\mu_3 = \lambda}$$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$\boxed{\mu_4 = 3\lambda^2 + \lambda}$

Moment Generating function about $\bar{x}$ (mean):

$$M_X(t) \text{ about mean} = E\left[e^{t(X-\bar{x})}\right]$$
$$= E\left[e^{t(X-\lambda)}\right]$$
$$= e^{-\lambda t} E\left[e^{tX}\right]$$
$$= e^{-\lambda t} M_X(t) \text{ about origin}$$
$$= e^{-\lambda t} e^{\lambda(e^t-1)} = e^{\lambda(e^t-1-t)}$$

$$\boxed{M_X(t) = e^{\lambda(e^t-1-t)}}$$

Moments about mean can be calculated by MGF about $\bar{x}$

$$\mu_4 = \left[\frac{d}{dt} M_X(t) \text{ (about mean)}\right]_{t=0} = 0$$

and so on.

Recurrence Relation for the central moments of Poisson Distribution

we have r^{th} moment about mean

$$\mu_r = E\{(X-\bar{x})^r\} = \sum_i P_i (x_i - \bar{x})^r$$
$$= \sum_{x=0}^{\infty} (x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{—— (1)}$$

differentiate (1) w.r.to λ, we get,

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} (-r)(x-\lambda)^{r-1} \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!}\left(-e^{-\lambda}\lambda^x + x\lambda^{x-1}e^{-\lambda}\right)$$

$$= (-r)\sum_{x=0}^{\infty} (x-\lambda)^{r-1} \cdot \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!} \cdot e^{-\lambda}\lambda^x\left(-1 + \frac{x}{\lambda}\right)$$

$$= -r\sum_{x=0}^{\infty} (x-\lambda)^{r-1} P(x) + \frac{1}{\lambda}\sum_{x=0}^{\infty} (x-\lambda)^{r+1} P(x)$$

$$\frac{d\mu_r}{d\lambda} = -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

$$\boxed{\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}}$$

Karl Pearson's Coefficient of Poisson distribution :

$$\boxed{\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda}}$$

$$\boxed{\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}}$$

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\lambda^2+\lambda}{\lambda^2} = 3+\frac{1}{\lambda}}$$

$$\boxed{\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}}$$

Distribution Function of Poisson Distribution:

$$F(x) = P(X \leq x) = \sum_{i=0}^{x} P(X=i)$$

or $F(x) = \sum_{r=0}^{x} P(X=r)$

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

Probability Generating Function of Poisson distribution:

$$G_X(z) = \sum_{r=0}^{\infty} \frac{e^{-\lambda}\lambda^r}{r!} z^r$$

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

$$\boxed{G_X(z) = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}}$$

Mode of Poisson Distribution:

Value of r for which $P(X=r)$ is maximum.

Now $P(r) > P(r-1)$ and $P(r) > P(r+1)$

Now $\frac{P(r)}{P(r-1)} > 1$

$$\Rightarrow \frac{\frac{e^{-\lambda}\lambda^r}{r!}}{\frac{e^{-\lambda}\lambda^{r-1}}{(r-1)!}} > 1 \quad \Rightarrow \frac{\lambda}{r} > 1 \text{ or } \lambda > r \quad \text{— (1)}$$

and $\frac{P(r)}{P(r+1)} > 1 \Rightarrow \frac{(r+1)}{\lambda} > 1$ or $(r+1) > \lambda$ — (2).

or $r > (\lambda - 1)$

From (1) and (2)

$$\boxed{\lambda-1 < r < \lambda}$$

Curve Fitting

In Applied Mathematics, many times it is required to express a given data (obtained from observations) in the form of a law connecting the variables involved. Such a law inferred by scheme is known as empirical law.

Several equations of diffrent types can be obtained to express the given date approximately. The process of finding such an eq^n of 'best fit' is known as curve-fitting. The best method of curve fitting is least square method.

Fitting a straight line

Here we use principle of least squares which states that the sum of squares of errors of estimation should be minimum.

If we want to fit a straight line $y = a + bx$ to the data given with n pts. $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$

then

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Normal eq^n are $\frac{\partial S}{\partial a} = 0 \qquad \frac{\partial S}{\partial b} = 0$

$$\Rightarrow -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

and $$-2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0$$

or $$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i \quad , \quad \sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2$$

Fitting a Parabola

we have to fit a parabola $y = a + bx + cx^2$
to the data of given n pts. $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$

$$S = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)^2$$

Normal eq^{n}s are

$\frac{\partial S}{\partial a} = 0$, $\frac{\partial S}{\partial b} = 0$, $\frac{\partial S}{\partial c} = 0$ which gives on simplification

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3$$

$$\sum_{i=1}^{n} x_i^2 y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4$$

Further by simplifying these simultaneous eq^{n}s we get values of a, b and c and required values of parabola.

Q.1 Fit a straight line to the following data

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

Sol^n Let the line to be fitted is $y = a + bx$

By the principle of least square the normal eq^{n}s are

$\Sigma y = na + b\Sigma x$, $\Sigma xy = a\Sigma x + b\Sigma x^2$

x	y	x^2	xy
1	2.4	1	2.4
2	3	4	6.0
3	3.6	9	10.8
4	4	16	16.0
6	5	36	30.0
8	6	64	48.0

$\Sigma x = 24$
$\Sigma y = 24$
$\Sigma x^2 = 130$
$\Sigma xy = 113.2$

substituting these values in eq^n we get

$$24 = 6a + 24b$$

$$113.2 = 24a + 130b$$

$$\Rightarrow b = \frac{17.2}{34} = 0.506 \quad , a = 4 \cdot 1.976$$

$$y = 1.976 + 0.506x$$

Q.2 Find the least square fit of the form $y = a_0 + a_1 x^2$ to the following date

x	-1	0	1	2
y	2	5	3	0

Sol^n put $x^2 = X$ we have, $y = a_0 + a_1 X$

The normal eq^n's are

$$\Sigma y = 4a + a_1 \Sigma X, \quad \Sigma Xy = a_0 \Sigma X + a_1 \Sigma X^2$$

x	y	X	X^2	Xy
-1	2	1	1	2
0	5	0	0	0
1	3	1	1	3
2	0	4	16	0
	$\Sigma y = 10$	$\Sigma X = 6$	$\Sigma X^2 = 18$	$\Sigma Xy = 5$

$\Rightarrow$ eq^n becomes

$$10 = 4a_0 + 6a_1 \quad , \quad 5 = 6a_0 + 18a_1$$

On solving $a_0 = 4.167, \quad a_1 = -1.111$

Hence the curve of best fit is

$$y = 4.167 - 1.111X \quad \text{or} \quad 4.167 - 1.111x^2$$

4 12

Q.3 Find the best values of x, y and z satisfying the following eq^{n}s.

$$x + 2y + z = 1$$
$$2x + y + z = 4$$
$$-x + y + 2z = 4$$
$$4x + 2y - 5z = -7$$

Sol^n In order to obtain the best values of x, y and z the normal eq^{n}s are

$$\frac{\partial S}{\partial x} = 0, \quad \frac{\partial S}{\partial y} = 0, \quad \frac{\partial S}{\partial z} = 0$$

$$S = (x+2y+z-1)^2 + (2x+y+z-4)^2 + (-x+y+2z-4)^2 + (4x+2y-5z+7)^2$$

$$\frac{\partial S}{\partial x} = 0 \Rightarrow 2(x+2y+z-1) + 4(2x+y+z-4) - 2(-x+y+2z-4) + 8(4x+2y-5z+7) = 0$$

$$\Rightarrow 40x + 22y - 38z = -46 \quad -(1)$$

$$\frac{\partial S}{\partial y} = 0 \Rightarrow 4(x+2y+z-1) + 2(2x+y+z-4) + 2(-x+y+2z-4) + 4(4x+2y-5z+7) = 0$$

$$\Rightarrow 22x + 20y - 10z = -8 \quad -(2)$$

$$\frac{\partial S}{\partial z} = 0 \Rightarrow 2(x+2y+z-1) + 2(2x+y+z-4) + 4(-x+y+2z-4) - 10(4x+2y-5z+7) = 0$$

$$\Rightarrow -38x - 10y + 62z = 96 \quad -(3)$$

On solving (1), (2) & (3) we get

$$x = 1.16, \quad y = -0.76, \quad z = 2.8$$

Scanned by CamScanner

Q.4 Fit a second degree parabola to the following data 513

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

Sol^n Let eq^n of parabola is

$$y = a + bx + cx^2 \quad , n = 9$$

Normal eq^{n}s are

$$\Sigma y = na + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

x	y	x^2	x^3	x^4	xy	x^2y
1	2	1	1	1	2	2
2	6	4	8	16	12	24
3	7	9	27	81	21	63
4	8	16	64	256	32	128
5	10	25	125	625	50	250
6	11	36	216	1296	6	396
7	11	49	343	2401	77	539
8	10	64	512	4096	80	640
9	9	81	729	6561	81	729
Σ 45	74	285	2025	15333	421	2771

Eq^n becomes

$$9a + 45b + 285c = 74$$
$$45a + 285b + 2025c = 421$$
$$285a + 2025b + 15333c = 2771$$

or $a + 5b + 31.67c = 8.22$

$a + 6.33b + 45c = 9.35$

$a + 7.10b + 53.8c = 9.72$

$\Rightarrow c = -0.26, \; b = 3.45, \; a = -0.78$

$y = -0.78 + (3.45)x - 0.26x^2$

Q.5 Fit a straight line for the following data

x	1	2	3	4	5
y	35	65	100	138	170

$y = 0.2 + 34x$

CORRELATION & REGRESSION

Correlation

The correlation coefficient tells us how strongly two variables are related, but it does not give us the magnitude of change of one variable due to other variable.

Ex. crime rate & unemployment rate

Karl Pearson coefficient of Correlation

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

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

$$= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

σ_x = Standard deviation of variable x

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

σ_y = Standard deviation of variable y

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$\bar{x}$ = Mean of variable x $= \frac{1}{n}\sum_{i=1}^{n} x_i$

$\bar{y}$ = Mean of variable y $= \frac{1}{n}\sum_{i=1}^{n} y_i$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{\cancel{n}}\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

or

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2}}$$

If values of x, y are very big then we can calculate

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i v_i - \frac{1}{n}\sum_{i=1}^{n} u_i \sum_{i=1}^{n} v_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} u_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} u_i\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n} v_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} v_i\right)^2}}$$

$$\frac{\sum uv - \frac{\sum u \sum v}{n}}{\sqrt{\sum u^2 - \frac{(\sum u)^2}{n}}\sqrt{\sum v^2 - \frac{(\sum v)^2}{n}}}$$

where $u_i = \frac{x_i - a}{h}$ $\quad v_i = \frac{y_i - b}{k}$

a, b are means of data ~~of~~ variables x and y.

h, k are class intervals of data x and y

$-1 \leq r \leq 1$

1. If $r = 0 \Rightarrow$ variables are unrelated
2. If $r = 1 \Rightarrow$ perfect and positive correlation
3. If $r = -1 \Rightarrow$ Perfect and negative correlation
4. $0 < r < 1 \Rightarrow$ positive correlation
5. $-1 < r < 0 \Rightarrow$ negative correlation.

Q.1 Calculate the correlation coefficient for the following heights 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

Solⁿ The correlation coefficient is given by

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\sum xy - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}}$$

x	y	x^2	y^2	xy
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
544	552	37028	38132	37560

$$\bar{x} = \frac{\sum x}{n} = \frac{544}{8} = 68$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1}{8}552 = 69$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2} = \sqrt{\frac{37028}{8} - (68)^2} = \sqrt{4.5} = 2.121$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2} = \sqrt{\frac{38132}{8} - (69)^2} = \sqrt{5.5} = 2.345$$

$$Cov(x,y) = \frac{1}{n}\sum xy - \bar{x}\bar{y} = \frac{1}{8}37560 - 68 \times 69 = 3$$

$$r = \frac{3}{2.121 \times 2.345} = 0.6032$$

Aliter Way

x	y	$u=x-68$	$v=y-69$	u^2	v^2	uv
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
		0	0	36	44	24

Now $\bar{u}=0$ $\quad \bar{v}=0$

$$\text{cov}(u,v) = \frac{1}{n}\Sigma uv - \bar{u}\bar{v} = \frac{1}{8}\times 24 = 3$$

$$\sigma u = \sqrt{\frac{1}{n}\Sigma u^2 - \bar{u}^2} = \sqrt{\frac{1}{8}36} = \sqrt{4.5} = 2.121$$

$$\sigma v = \sqrt{\frac{1}{n}\Sigma v^2 - \bar{v}^2} = \sqrt{\frac{1}{8}\times 44} = \sqrt{5.5} = 2.345$$

$$r = \frac{\text{cov}(u,v)}{\sigma u\,\sigma v} = \frac{3}{2.121\times 2.345} = 0.6032$$

Q.2 Given

Coefficient of correlation $r = 0.8$

Standard deviation of y series $= 2.5$

Product of deviations taken from their respective airthmetic means $= 60$

Sum of squares of deviations taken from airthmetic means of x series $= 90$.

Find the number of items.

Scanned by CamScanner

Sol^n We have given

$r = 0.8$, $\Sigma y = 2.5$, $\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 60$

$\Sigma(x_i - \bar{x})^2 = 90$

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2}\sqrt{\Sigma(y_i - \bar{y})^2}}$$

$$= \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$$

$\sigma_x = \sqrt{\frac{1}{n}\Sigma(x_i - \bar{x})^2}$, $\sigma_y = \sqrt{\frac{1}{n}\Sigma(y_i - \bar{y})^2}$

$$r = 0.8 = \frac{60}{n\sqrt{\frac{90}{n}} \times 2.5} \Rightarrow 0.8 = \frac{60}{\sqrt{n}\sqrt{90} \times 2.5}$$

$$\text{or } \sqrt{n} = \frac{60}{2.5 \times 0.8 \times \sqrt{90}}$$

$$= 9.999 \simeq 10$$

Q.3 Calculate the coefficient of correlation bet^n x & y using the following data

Let $a = 7$, $b = 15$

x	y	$u = x-7$	$v = y-15$	u^2	v^2	uv
1	8	-6	-7	36	49	42
3	12	-4	-3	16	9	12
5	15	-2	0	4	0	0
7	17	0	2	0	4	0
8	18	1	3	1	9	3
10	20	3	5	9	25	15
		$\Sigma u = -8$	0	66	96	72

$\bar{u} = \frac{\Sigma u}{n} = \frac{-8}{6} = \frac{-4}{3}$ $\qquad \bar{v} = \frac{\Sigma v}{n} = 0$

$$\sigma_x^2 = \frac{\Sigma u^2}{n} - (\bar{u})^2 = \frac{66}{6} - \frac{16}{9} = \frac{83}{9}$$

$$\sigma_y^2 = \frac{\Sigma v^2}{n} - (\bar{v})^2 = \frac{96}{6} - 0 = 16$$

$$\text{Cov}(x,y) = \frac{\Sigma uv}{n} - \bar{u}\bar{v} = \frac{72}{6} - 0 = 16$$

$$r = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{12}{4\sqrt{\frac{83}{9}} \times 4} = \frac{9}{9.11}$$

$$= 0.988$$

Q4 Calculate the coefficient of correlation

x	y	$u = x-5$	$v = y-12$	u^2	v^2	uv
1	9	−4	−3	16	9	12
2	8	−3	−4	9	16	12
3	10	−2	−2	4	4	4
4	12	−1	0	1	0	0
5	11	0	−1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	16	9	12
		0	0	60	60	57

$\bar{u} = \bar{v} = 0$ $\quad$ $\text{Cov}(u,v) = \frac{1}{n} \Sigma uv - \bar{u}\bar{v}$

$$= \frac{57}{9} = 6.333$$

$$\sigma_u = \sqrt{\frac{1}{n} \Sigma u^2 - \bar{u}^2} = \sqrt{\frac{60}{9}} = \sqrt{6.666} = 2.582 = \sigma_v$$

$$r = \frac{\text{Cov}(u,v)}{\sigma_u \sigma_v} = \frac{6.333}{6.666} = 0.95$$

Rank correlation

Whenever it is not possible to measure any characteristic attribute like honesty, morality, beauty etc. then we assign rank of that attribute and then calculate the correlation coefficients.

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where $d_i = x_i - y_i$, $\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} [(x_i-\bar{x})^2 - (y_i-\bar{y})]^2$

Q.1 Obtain the rank correlation coefficient for the following data

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	74

Solⁿ.

X	68	64	75	50	64	80	75	40	55	64	
Rank(x_i)	4	6	2.5	9	6	1	2.5	10	8	6	
Y	62	58	68	45	81	60	68	48	50	70	
Rank (y_i)	5	7	3.5	10	1	6	3.5	9	8	2	
$d_i = x_i - y_i$	-1	-1	-1	-1	5	-5	-1	1	0	4	⇒ 0
d_i^2	1	1	1	1	25	25	1	1	0	16	⇒ 72

2.5 is repeated twice so it correlation factor

$(2.5)\ X = \frac{2(2^2-1)}{12} = \frac{1}{2}$ $\qquad$ X (for 6) $= \frac{3(3^2-1)}{12} = 2$

C.f. for $X = \frac{1}{2} + 2 = 5/2$

C.f. for $Y = \frac{2(2^2-1)}{12} = \frac{1}{2}$

The rank correlation coefficient

$$r = 1 - \frac{6\left[\sum d^2 + \frac{5}{2} + \frac{1}{2}\right]}{n(n^2-1)}$$

$$= 1 - \frac{6(72+3)}{10(100-1)} = 1 - \frac{6 \times 75}{10 \times 99} = 1 - \frac{5}{11} = \frac{6}{11}$$

$r = 0.545$

Q. 2 Ten competitors in a beality contest are ranked by three judges in the following order

I judge R_1	II judge R_2	III judge R_3	$D_1 = R_1 - R_2$	D_1^2	$D_2 = R_2 - R_3$	D_2^2	$D_3 = R_1 - R_3$	D_3^2
1	3	6	-2	4	-3	9	-5	25
6	5	4	1	1	1	1	2	4
5	8	9	-3	9	-1	1	-4	16
10	4	8	6	36	-4	16	2	4
3	7	1	-4	16	6	36	2	4
2	10	2	-8	64	8	64	0	0
4	2	3	2	4	-1	1	1	1
9	1	10	8	64	-9	81	-1	1
7	6	5	1	1	1	1	2	4
8	9	7	-1	1	2	4	1	1
				200		214		60

Rank correlation coefficient betn I & II judges

$$r = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6 \times 200}{10(100-1)} = -0.212$$

Rank co. coeff. betn II & III

$$r = 1 - \frac{6\sum D_2^2}{n(n^2-1)} = 1 - \frac{6(214)}{10(100-1)} = -0.297$$

Rank co. coeff. betn I & III

$$r = 1 - \frac{6\sum D_3^2}{n(n^2-1)} = 1 - \frac{6(60)}{10(100-1)} = 0.636$$

Q.3 The marks of 8 candidates in Mathematics and English are given as

Maths	76	90	98	69	54	82	67	52
English	25	37	56	12	7	36	23	11

Calculate rank correlation.

Solⁿ

x (Maths)	y (Eng.)	R_1	R_2	$D = R_1 - R_2$	D^2
76	25	4	4	0	0
90	37	2	2	0	0
98	56	1	1	0	0
69	12	5	6	-1	1
54	7	7	8	-1	1
82	36	3	3	0	0
67	23	6	5	1	1
52	11	8	7	1	1
					$\Sigma D^2 = 4$

$$r = 1 - \frac{6\Sigma D^2}{N(N^2-1)} = 1 - \frac{6 \times 4}{8(64-1)} = 0.925$$

Regression

Regression model help us to evaluate the magnitude of change in one variable due to other variable.

Equation of line of regression of y on x is

$$y - \bar{y} = \underbrace{\frac{r\sigma_y}{\sigma_x}}_{b_{yx}} (x - \bar{x})$$

$\rightarrow b_{yx}$ Regression coefficients y on x

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

The line of regression of x on y is

$$x-\bar{x} = \frac{cov(x,y)}{\sigma_y^2}(y-\bar{y})$$

or $$x-\bar{x} = \underbrace{\frac{r\sigma_x}{\sigma_y}}_{b_{xy}}(y-\bar{y})$$

Note:–

1. $\bar{x}, \bar{y}$ is point of intersection of the two lines of regression
2. $b_{yx} = \frac{r\sigma_y}{\sigma_x}$ and $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ is known as coefficient of regression.
3. If $r=0$ then $y=\bar{y}$ & $x=\bar{x}$ be two lines of regression
4. If $r=\pm 1$ then line of regression is

$$\frac{y-\bar{y}}{\sigma_y} = \pm\left(\frac{x-\bar{x}}{\sigma_x}\right)$$

Properties of Regression Coefficients

1. Correlation coefficient (r) is geometric mean betn b_{yx} & b_{xy}.
2. If one of the regression coefficient is greater than 1, the other must be less than unity.
3. Arithmatic mean of regression coefficients is greater than the correlation coefficients
4. Regression coefficients are independent of the change of origin but not of scale.

Q.1 Show that θ, the acute angle betn the two lines of regression is

$$\tan\theta = \left(\frac{1-r^2}{r}\right) \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}$$ Also Interpret the case when $r=0, \pm 1$.

Soln

The regression line of y on x is

$$(y-\bar{y}) = \frac{r\sigma_y}{\sigma_x}(x-\bar{x}) \qquad m_1 = \frac{r\sigma_y}{\sigma_x}$$

regression line of x on y is

$$x-\bar{x} = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \qquad m_2 = \frac{\sigma_y}{r\sigma_x}$$

$$\tan\theta = \pm\frac{(m_1-m_2)}{1+m_1 m_2}$$

$$\tan\theta = \pm\frac{\left(\frac{r\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}\right)}{\left(1+\frac{r\sigma_y}{\sigma_x}\cdot\frac{\sigma_y}{r\sigma_x}\right)} = \pm\frac{(r^2\sigma_x\sigma_y - \sigma_y\sigma_x)}{r\sigma_x^2+r\sigma_y^2}$$

$$= \pm\frac{\sigma_x\sigma_y(r^2-1)}{r(\sigma_x^2+\sigma_y^2)}$$

on positive sign $\tan\theta = \dfrac{\sigma_x\sigma_y(r^2-1)}{r(\sigma_x^2+\sigma_y^2)} = \dfrac{-(1-r^2)\sigma_x\sigma_y}{r(\sigma_x^2+\sigma_y^2)}$

(θ will be an obtuse angle)

Taking negative sign

$$\tan\theta = \frac{\sigma_x\sigma_y(1-r^2)}{r(\sigma_x^2+\sigma_y^2)}$$ (θ will be an acute angle)

when $r=0$, $\tan\theta = \infty$

$\Rightarrow \theta = 90^\circ \Rightarrow$ both the regression lines are $\perp$ to each other when $r = \pm 1$, $\tan\theta = 0$ or π

$\Rightarrow$ both the regression lines coincide each other and there is perfect correlation.

Q.2 Calculate the coefficient of correlation and obtain the line of regression for the following data

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

$\rightarrow$ Aliter

Sol^n As done earlier

$$r_{xy} = r_{uv} = 0.95$$

As $u = x-5$, $v = y-12$

$\bar{u} = \bar{x} - 5$ $\quad \bar{v} = \bar{y} - 12$

$\sigma_u^2 = \sigma_x^2$, $\sigma_v^2 = \sigma_y^2$

Line of regression of y on x is

$$y - \bar{y} = r_{xy}\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$y - (\bar{v}+12) = r_{uv}.\frac{\sigma_u}{\sigma_v}[x - (\bar{u}+5)]$$

$$y - (0+12) = 0.95\frac{\sqrt{60/9}}{\sqrt{60/9}}[x - (0+5)]$$

$$y - 12 = 0.95(x-5) \Rightarrow y = 0.95x + 7.25$$

Line of regression of x on y is

$$x - \bar{x} = r_{xy}\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$x - (\bar{u}+5) = r_{uv}[y - (\bar{v}+12)]$$

$$x - 5 = 0.95(y-12)$$

$$x = 0.95y - 6.4$$

Scanned by CamScanner

Q.3. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible. 13.

Variance of $x = 9$

Regression eqⁿ $8x - 10y + 66 = 0$, $40x - 18y = 214$

Find a) the mean values of x & y

Since both the lines of regression pass through $\bar{x}, \bar{y}$.

Hence

$$8\bar{x} - 10\bar{y} + 66 = 0$$

$$40\bar{x} - 18\bar{y} = 214 \longrightarrow \bar{x} = 13, \ \bar{y} = 17$$

on solving we get $\cancel{b_{xy} = \frac{10}{8}}$

$\cancel{b_{yx} = \frac{40}{18}}$

b) The standard deviation of y

Let assume the regression eqⁿ of x on y be

$$8x = -66 + 10y$$

$$x = -\frac{66}{8} + \frac{10}{8}y \Rightarrow b_{xy} = \frac{10}{8}$$

from second eqⁿ

$$-18y = -40x + 214$$

$$y = -\frac{214}{18} + \frac{40}{18}x \Rightarrow b_{yx} = \frac{40}{18}$$

Since both the regression coefficients are greater than one, our assumption is wrong, Hence the first eqⁿ is y on x.

Again

$$10y = 8x + 66$$

$$y = \frac{8}{10}x + 6.6 \Rightarrow b_{yx} = \frac{8}{10}$$

$$x = \frac{18}{40}y + \frac{214}{40} \Rightarrow b_{xy} = \frac{18}{40}$$

$$r = \pm\sqrt{b_{xy} \times b_{yx}} = \sqrt{0.6} = \sqrt{0.36} = \pm 0.6$$

c) Coefficient of correlation betⁿ x and y

Given $\sigma_x = 3$

$$b_{xy} = r\frac{\sigma_x}{\sigma_y} \Rightarrow \frac{18}{40} = 0.6 \times \frac{3}{\sigma_y} \Rightarrow \sigma_y = 4$$

Q.4 For a bivariate distribution $n = 18$, $\Sigma x^2 = 60$, $\Sigma y^2 = 96$ $\Sigma x = 12$, $\Sigma y = 18$, $\Sigma xy = 48$. Find the equations of the lines of regression and r.

Solⁿ $\bar{x} = \frac{\Sigma x}{n} = \frac{12}{18} = 0.667.$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{18}{18} = 1$$

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{60}{18} - (0.667)^2 = 2.8884$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{96}{18} - 1 = 4.3333$$

$$\text{cov}(x, y) = \frac{\Sigma xy}{n} - \bar{x}\bar{y} = \frac{48}{18} - (0.667)(1) = 1.9997$$

Line of regression of y on x is

$$y - \bar{y} = \frac{\text{cov}(x, y)}{\sigma_x^2}(x - \bar{x})$$

$$y - 1 = \frac{1.9997}{2.888}(x - 0.667)$$

$$= 0.692(x - 0.667)$$

$$\Rightarrow 0.692x - y + 0.538 = 0$$

$$\Rightarrow y = 0.692x + 0.538$$

Line of regression of x on y is

$$x - \bar{x} = \frac{\text{cov}(x, y)}{\sigma_y^2}(y - \bar{y}) \Rightarrow x - 0.667 = \frac{1.9997}{4.333}(y - 1)$$

$$x - 0.667 = 0.4615(y - 1)$$

$$x = 0.4615y + 0.2055$$

Also $r^2 = b_{xy} \cdot b_{yx} = 0.4615 \times 0.692 = 0.3194$

$$r = 0.57$$