16/04/19

M:2B SPECIAL FUNCTION

Sol. of L.E in cylinderical system
$\downarrow$
Series sol. of Bessel's Differential equation
$\downarrow$
Bessel's function

Sol^n of LE in spherical system
$\downarrow$
Series solution of legendre's DE
$\downarrow$
Legendre Polynomial

special function

**

Bessel's differential equation

Series solution of BDE leading to Bessel's solution

Bessel's differential equation is given by

$$x^2y'' + xy' + (x^2 - n^2)y = 0 \quad \text{——} (1)$$

$\therefore$ $P_0(x)$ = Coeff of $y'' = x^2 = 0$ at $(x=0)$

Let $y = \sum_{r=0}^{\infty} a_r x^{k+r}$, $a_0 \neq 0$ —— (2) be the series solution of (1)

differentiating (2) w.r.t x we get;

$$y' = \sum_{r=0}^{\infty} (k+r)a_r x^{k+r-1} \quad \text{——} (3)$$

differentiating (3) w.r.t x we get;

$$y'' = \sum_{r=0}^{\infty} (k+r)(k+r-1) \cdot a_r x^{k+r-2}$$

$$y'' = \sum_{r=0}^{\infty} \left[(k+r)^2 - (k+r)\right] a_r \cdot x^{k+r-2} \quad \text{——} (4)$$

Substituting the values of y, y'', y' obtained from eq^n (2), (3) and (4) in eq^n (1).

$$x^2 \sum_{r=0}^{\infty} \left[(k+r)^2 - (k+r)\right] a_r x^{k+r-2} + x \sum_{r=0}^{\infty} (k+r) a_r x^{k+r-1} + (x^2 - n^2) \sum_{r=0}^{\infty} a_r x^{k+r} = 0$$

$$\Rightarrow \sum_{r=0}^{\infty} \left[(k+r)^2 - (k+r)\right] a_r x^{k+r} + \sum_{r=0}^{\infty} (k+r) a_r x^{k+r} + \sum_{r=0}^{\infty} (-n^2) a_r x^{k+r} + \sum_{r=0}^{\infty} a_r x^{k+r+2} = 0.$$

$$\sum_{r=0}^{\infty} \left[(k+r)^2 - (k+r) + (k+r) - n^2\right] a_r \cdot x^{k+r} + \sum_{r=0}^{\infty} a_r \cdot x^{k+r+2} = 0$$

$$\underbrace{\sum_{r=0}^{\infty} \left[(k+r)^2 - n^2\right] a_r x^{k+r}}_{\mathrm{I}} + \underbrace{\sum_{r=0}^{\infty} a_r x^{k+r+2}}_{\mathrm{II}} = 0 \quad \text{——} (5)$$

On equating the coefficient of x^k to zero in eq^n (5).
ie by putting $r=0$ in I, we get;

$(k^2 - n^2)a_0 = 0$, $k^2 - n^2 = 0$ as $a_0 \neq 0$

$k^2 = n^2$, $k = \pm n$

Now by putting $r=1$ in I, we get;

$[(k+1)^2 - n^2]a_1 = 0.$

$\Rightarrow$ $a_1 = 0$ as $(k+1)^2 - n^2 \neq 0$

since $k = \pm n$

On equating the coefficient of x^{k+r} to zero in eq^n (5)
ie by putting $r=r$ in (I) and $r = r-2$ in II we get.

$[(k+r)^2 - n^2] a_r + a_{r-2} = 0.$

$[(k+r)^2 - n^2] a_r = -a_{r-2}$

$$\boxed{a_r = (-a_{r-2})/((k+r)^2 - n^2)} \quad , r \geq 2.$$

Put $k = n$ in eq^n (6)

$$a_r = \frac{-a_{r-2}}{[(n+r)^2 - n^2]} = \frac{-a_{r-2}}{[\cancel{n^2} + 2nr + r^2 - \cancel{n^2}]}$$

$$\boxed{a_r = \frac{-a_{r-2}}{(2n+r)\cdot r}, \quad r \geq 2} \quad \text{—— (7)}$$

Put $r = 2$ in (7)

$$a_2 = \frac{-a_0}{(2n+2)2} = \frac{-a_0}{4(n+1)}$$

Put $r = 3$ in (7)

$$a_3 = \frac{-a_1}{(2n+3)3} = 0 \quad \text{as } (a_1 = 0)$$

Put $r = 4$ in eq^n (7)

$$a_4 = \frac{-a_2}{(2n+4)4} = \frac{-a_2}{8(n+2)}$$

$$a_4 = \frac{a_0}{32(n+1)(n+2)}$$

Let $y = y_1$ for $k = n$ be the solution of (1)

$\therefore$ (2) becomes

$$y_1 = \sum_{r=0}^{\infty} a_r \cdot x^{n+r}$$

$$y_1 = x^n \sum_{r=0}^{\infty} a_r \cdot x^r$$

$$y_1 = x^n[a_0 + \cancel{a_1 x}^{\,0} + a_2x^2 + \cancel{a_3x^3}^{\,0} + a_4x^4 + \dots]$$

$$y_1 = x^n a_0\left[1 - \frac{x^2}{4(n+1)} + \frac{x^4}{32(n+1)(n+2)} - \dots\right] \quad \text{—— (8)}$$

Let $y = y_2$ (for $k = -n$) be the series solution of (1)

$$\therefore y_2 = a_0 x^{-n}\left[1 - \frac{x^2}{4(-n+1)} + \frac{x^4}{32(-n+1)(-n+2)} - \dots\right] \quad \text{—— (9)}$$

(from (8) by replacing n by $-n$)

Complete solution of (1) is given by ;

$y = Ay_1 + By_2$, where A and B are arbitary constant and y_1 and y_2 are given by (8) and (9).

Let $a_0 = \dfrac{1}{2^n \Gamma(n+1)}$

$\therefore$ (8) becomes

$$y_1 = \frac{1}{2^n\Gamma(n+1)} x^n\left[1 - \frac{x^2}{4(n+1)} + \frac{x^4}{32(n+1)(n+2)} - \dots\right]$$

$$y_1 = \left[\left(\frac{x}{2}\right)^n \frac{1}{\Gamma(n+1)} - \left(\frac{x}{2}\right)^{n+2} \frac{1}{(n+1)\Gamma(n+1)} + \left(\frac{x}{2}\right)^{n+4} \frac{1}{2(n+1)(n+2)\Gamma(n+1)} - \dots\right]$$

$$y_1 = \left[\left(\frac{x}{2}\right)^n \frac{1}{\Gamma(n+1)\,0!} - \left(\frac{x}{2}\right)^{n+2} \frac{1}{\Gamma(n+2)\,1!} + \left(\frac{x}{2}\right)^{n+4} \frac{1}{\Gamma(n+3)\,2!} - \dots\right]$$

$$y_1 = \sum_{r=0}^{\infty} \frac{(-1)^r}{(r!)\Gamma(n+r+1)} \left(\frac{x}{2}\right)^{n+2r} \qquad \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(n+r+1)} \left(\frac{x}{2}\right)^{n+2r}$$

Which is called the Bessel's function of first kind of order n and is given by above denoted by $J_n(x)$

$$J_n(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(n+r+1)} \left(\frac{x}{2}\right)^{n+2r}$$

The solution for $k = -n$ is denoted by $J_{-n}(x)$.

General solution of Bessel's eq^n (1) is given by

$y = a\,J_n(x) + b\,J_{-n}(x)$ where a and b are arbitary constants.

① P.T. $J_n(x) = \sqrt{\frac{2}{\pi x}} \sin x$ for $n = 1/2$ and $J_n(x) = \sqrt{\frac{2}{\pi x}} \cos x$ for $n = -1/2$

***

$$\sin x = x - \frac{x^3}{\lfloor 3} + \frac{x^5}{\lfloor 5} - \dots$$

$$\cos x = 1 - \frac{x^2}{\lfloor 2} + \frac{x^4}{\lfloor 4} - \dots$$

$$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(3/2) = \frac{1}{2}\sqrt{\pi}, \quad \Gamma(5/2) = \frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}$$

Proof: $$J_n(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(n+r+1)} \left(\frac{x}{2}\right)^{n+2r} \quad \text{—— ①}$$

Put $n = \frac{1}{2}$ in ①

$$J_{1/2}(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma\left(\frac{1}{2}+r+1\right)} \left(\frac{x}{2}\right)^{1/2 + 2r + \cancel{1/2} - \cancel{1/2}}$$

$$= \left(\frac{x}{2}\right)^{-1/2} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(r+3/2)} \left(\frac{x}{2}\right)^{1+2r}$$

$$= \left(\frac{2}{x}\right)^{1/2} \left[\frac{(-1)^0}{0!\,\Gamma(3/2)} \left(\frac{x}{2}\right) + \frac{(-1)}{1!\,\Gamma(5/2)} \left(\frac{x}{2}\right)^3 + \frac{(-1)^2}{2!\,\Gamma(7/2)} \left(\frac{x}{2}\right)^5 + \dots \right]$$

$$= \sqrt{\frac{2}{x}} \left[\frac{1}{\cancel{1/2}\sqrt{\pi}} \left(\frac{x}{\cancel{2}}\right) - \frac{1}{\frac{3}{\cancel{2}}\cdot\frac{1}{\cancel{2}}\sqrt{\pi}} \frac{x^3}{\cancel{8}_2} + \frac{1}{2\cdot\frac{5}{\cancel{2}}\cdot\frac{3}{\cancel{2}}\cdot\frac{1}{\cancel{2}}\sqrt{\pi}} \frac{x^5}{\cancel{32}\,4} + \dots \right]$$

$$= \sqrt{\frac{2}{\pi x}} \left[x - \frac{x^3}{\lfloor 3} + \frac{x^5}{\lfloor 5} - \dots \right]$$

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x.$$

Put $n = -1/2$ in ①

$$J_{-1/2}(x) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(-1/2 + r + 1)} \left(\frac{x}{2}\right)^{-1/2 + 2r}$$

$$= \left(\frac{x}{2}\right)^{-1/2} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!\,\Gamma(r+1/2)} \left(\frac{x}{2}\right)^{2r}$$

$$= \left(\frac{2}{x}\right)^{1/2} \left[\frac{(-1)^0}{0!\,\Gamma(1/2)} \left(\frac{x}{2}\right)^0 + \frac{(-1)^1}{\lfloor 1\,\Gamma(3/2)} \left(\frac{x}{2}\right)^2 + \frac{(-1)^2}{\lfloor 2\,\Gamma(5/2)} \left(\frac{x}{2}\right)^4 - \dots \right]$$

$$= \sqrt{\frac{2}{x}} \left[\frac{1}{\sqrt{\pi}} - \frac{x^2}{\cancel{1/2}\sqrt{\pi} * \cancel{4}_2} + \frac{x^4}{2 * \cancel{3/2}\,\cancel{1/2}\sqrt{\pi} * \cancel{16}\,4} - \dots \right]$$

$$= \sqrt{\frac{2}{\pi x}} \left[1 - \frac{x^2}{\lfloor 2} + \frac{x^4}{\lfloor 4} - \dots \right]$$

$$J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}} \cos x$$

Series solution of Legendre differential equation leading to Legendre polynomial

Equation (Differential) is given by

$$(1-x^2) y'' - 2xy' + n(n+1)y = 0 \quad \text{—— ①}$$

$P_0(x) = $ coefficient of $y'' = 1 - x^2 \neq 0$ at $x = 0$

Let $y = \sum_{r=0}^{\infty} a_r x^r$ —— ② be the series solution of ①

differentiating ② w.r.t x

$$y' = \sum_{r=0}^{\infty} r a_r x^{r-1} \quad \text{—— ③}$$

differentiating ③ w.r.t x we get;

$$y'' = \sum_{r=0}^{\infty} r(r-1) a_r x^{r-2} \quad \text{—— ④}$$

Substituting the values of y, y' and y'' in eq^n ①

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

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

$$\sum_{r=0}^{\infty} r(r-1)\, a_r\, x^{r-2} + \sum_{r=0}^{\infty}\left[-r^2 + \cancel{r} - 2\cancel{r} + n^2 + n\right] a_r\, x^r = 0.$$

$$\sum_{r=0}^{\infty} r(r-1) a_r x^{r-2} + \sum_{r=0}^{\infty}\left[(n-r)(n+r) + n - r\right] a_r x^r = 0.$$

$$\underbrace{\sum_{r=0}^{\infty} r(r-1)\, a_r\, x^{r-2}}_{\mathrm{I}} + \underbrace{\sum_{r=0}^{\infty} (n-r)(n+r+1)\, a_r\, x^r}_{\mathrm{II}} = 0$$

On equating the coefficients of x^{-2} to zero by substituting $r=0$ in (I)

$0(-1)\, a_0 = 0 \Rightarrow a_0 \neq 0.$

On equating the coefficient of x^{-1} to zero ie by putting $r = 01$ in I.

$1(0) \cdot a_1 = 0 \Rightarrow a_1 \neq 0.$

On equating the coefficient of x^r to zero ie by putting $r = r+2$ in I and $r = r$ in II

$$(r+2)(r+2-1).\, a_{r+2} + (n-r)(n+r+1)a_r = 0$$

$$(r+2)(r+1)\, a_{r+2} = -[(n-r)(n+r+1) a_r]$$

$$a_{r+2} = \frac{-(n-r)(n+r+1)\, a_r}{(r+1)\,(r+2)}, \quad r \geq 0 \quad \text{—— (5)}$$

Put $r=0$ in eq (5), $\quad a_2 = \dfrac{-n\,(n+1)\, a_0}{2}$

Put $r=1$ in eq^n (5)

$$a_3 = \frac{-(n-1)(n+2)\, a_1}{6}$$

$$a_4 = \frac{-(n-2)(n+3)}{12}\, a_2 = \frac{n(n+1)(n-2)(n+3)}{24}\, a_0$$

Put $r=3$ in eq^n (5), $\quad a_5 = \dfrac{-(n-3)(n+4)}{20}\, a_3 = \dfrac{(n-1)(n+2 \ldots}{}$

Put from (2), $\quad y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \ldots$

$$y = a_0 + a_1 x - \frac{n(n+1)}{2} a_0 x^2 - \frac{n(n-1)(n+2)}{6} a_1 x^3 + \frac{n(n+1)(n-2)(n+3)}{24} a_0 x^4 + \frac{(n-1)(n+2)(n-3)(n+\ldots}{120}$$

$$y = a_0 + a_1 x - \frac{n(n+1)}{\lfloor 2} a_0 x^2 - \frac{n(n-1)(n+2)}{\lfloor 3} a_1 x^3 + \frac{n(n+1)(n-2)(n+3)}{\lfloor 4} a_0 x^4 + \frac{(n-1)(n+2)(n-3)(n+4)}{\lfloor 5} a_1 \ldots$$

$$y = a_0\left[1 - \frac{n(n+1)}{\lfloor 2} x^2 + \frac{n(n+1)(n-2)(n+3)}{\lfloor 4} x^4 - \ldots\ldots\right]$$

$$+ a_1\left[x - \frac{(n-1)(n+2)x^3}{\lfloor 3} + \frac{(n-1)(n+2)(n-3)(n+4)}{\lfloor 5} x^5 \ldots\right.$$

$$y_0 = a_0\, u(x) + a_1 v(x) \quad \text{—— (6)}.$$

where $u(x)$ and $v(x)$ are infinite series

Eq^n (6) is the series solution of legendre differential eq^n (1).

If n is a positive even integer then $u(x)$ reduces to a … of degree 'n'; and if n is a positive odd integer then $v(x)$ reduces to a polynomial of degree 'n'.

Since the polynomial $u(x)$ and $v(x)$ contain alternate power… ∴ the general form of the polynomial that represents the e… two polynomial in descending powers of 'x' can be written a…

$$y = f(x) = a_n x^n + a_{n-2} x^{n-2} + a_{n-4} x^{n-4} + \ldots\ldots +$$

where $F(x) = \begin{cases} a_0, & \text{if } n \text{ is a +ve even integer} \ldots \end{cases}$

from (5) ; $a_r = -\dfrac{(r+1)(r+2)}{(n-r)(n+r+1)} a_{r+2}$ —— (8)

Put $r = n-2$ in (7)

$$a_{n-2} = \frac{-(n-2+1)(n-\cancel{2}+\cancel{2})}{[\cancel{n}-(\cancel{n}-2)][n+n-2+1]} a_{n-\cancel{2}+\cancel{2}}$$

$$a_{n-2} = \frac{-(n-1)n}{2\cdot(2n-1)} a_n$$

Put $r = n-4$ in eq^n (8)

$$a_{n-4} = \frac{-(n-4+1)(n-4+2)}{[n-(n-4)][n+n-4+1]} a_{n-4+2} \text{ ——}$$

$$a_{n-4} = -\frac{(n-3)(n-2)}{4(2n-3)} a_{n-2} = \frac{n(n-1)(n-2)(n-3)}{8(2n-1)(2n-3)} a_n$$

eq^n (7) becomes ; $y = f(x) \Rightarrow$

$$y = f(x) = a_n \left[x^n - \frac{n(n-1)}{2(2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{8(2n-1)(2n-3)} x^{n-4} + \dots G(x) \right]$$

where ; $G(x) = \begin{cases} \dfrac{a_0}{a_n} , & \text{if } n \text{ is a positive even integer} \\ \dfrac{a_1 x}{a_n} , & \text{if } n \text{ is a +ve odd integer} \end{cases}$

If the constant a_n is choosen that $y = f(x)$ becomes 1; when $x = 1$ then the polynomials so obtained are called legendre polynomials of degree 'n' denoted by $P_n(x)$.

Let $a_n = \dfrac{1 \cdot 3 \cdot 5 \dots (2n-1)}{\lfloor n}$ in order to meet the above said requirement.

ie $P_n(1) = 1$

$$\therefore P_n(x) = \frac{1 \cdot 3 \cdot 5 \dots (2n-1)}{\lfloor n} \left[x^n - \frac{n(n-1)}{2(2n-1)} x^{n-2} + \frac{n(n-1)(n-2)(n-3)}{8(2n-1)(2n-3)} * x^{n-4} - \dots \right] \text{ —— (9)}$$

Put $n = 0$ in (9)

$P_0(x) = \dfrac{1}{\lfloor 0} [x^0] = 1$

[polynomials never have negative Powers so rest of the series is neglected.]

Put $n = 1$ in (9)

$P_1(x) = \dfrac{1}{\lfloor 1} [x] = x$

Put $n = 2$ in (9)

$P_2(x) = \dfrac{1 \cdot 3}{\lfloor 2} \left[x^2 - \dfrac{2(1)\, x^0}{2 * 3} \right] = \dfrac{\cancel{3}}{2} \dfrac{(3x^2-1)}{\cancel{3}} = \dfrac{3x^2-1}{2}$

Put $n = 3$ in (9)

$P_3(x) = \dfrac{1 \cdot 3 \cdot 5}{\lfloor 3} \left[x^3 - \dfrac{3(2)}{2(5)} x \right] = \dfrac{\cancel{3} * \cancel{5}}{\cancel{3} * 2} \dfrac{[5x^3 - 3x]}{\cancel{5}} = \dfrac{5x^3 - 3x}{2}$

Put $n = 4$ in (9)

$P_4(x) = \dfrac{1 \cdot 3 \cdot 5 \cdot 7}{\lfloor 4} \left[x^4 - \dfrac{\overset{2}{\cancel{4}} * \cancel{3}}{\cancel{2} * 7} x^2 + \dfrac{\cancel{4} \times 3 \times \cancel{2} \times 1}{\cancel{8} \times 7 \times 5} x^0 \right]$

$= \dfrac{1 \cdot \cancel{3} \cdot \cancel{5} \cdot \cancel{7}}{4 \cdot \cancel{3} \cdot 2} \left[\dfrac{-30x^2 + 35x^4 + 3}{\cancel{35}} \right]$

$= \dfrac{35x^4 - 30x^2 + 3}{8}$

IMP

- Series soln of Bessel's differential eq^n.
- Series soln of legendre diff eq^n.
- Orthogonality of Bessel's function. If α and β are distinct roots of $J_n(\alpha x) = 0$ then

$$\int_0^1 x \cdot J_n(\alpha x) \cdot J_n(\beta x) \cdot dx = \begin{cases} 0 & \text{if } \beta \neq \alpha \\ \dfrac{[J_n'(\alpha)]^2}{2} = \dfrac{[J_{n+1}(\alpha)]^2}{2} & \text{if } \beta = \alpha . \end{cases}$$

- Derive rodrigue's formula for LF.

$P_n(x) = \dfrac{1}{2^n \lfloor n} \dfrac{d^n}{dx^n} (x^2-1)^n$ Hence obtain $P_0(x)$, $P_1(x)$,

$$\int_{-1} P_m(x)\, P_n(x) \cdot dx = \begin{cases} 0, & \text{if } m \neq n \\ \dfrac{2}{2n+1}, & \text{if } m = n \end{cases}$$

MODULE:05

23 April 19

Stochastic process or random process

Let X be a random variable then $X = X(s)$ $s \in S$.

Let $t \in T$ be a real parameter.

Let $X = X(s,t)$. Let this be denoted by $X(t)$ or X_t.

Then corresponding to each $t \in T$ there exists a random variable $X(t)$. The family of all such random variables $\{X(t) \mid t \in T\}$ is called random process or stochastic process with t as the parameter.

The set T over which t varies is called the <u>index set</u>.

For a specified value of 't' the value $X_t(s)$ assumed by $X(t)$ as 's' varies over 'S' are called the <u>state space of the process</u>.

In practical situation

Eg: Let a coin be tossed four times

Let $t = 1, 2, 3, 4 \in T$ Let the random variable $X(t)$

$$X(t) = \begin{cases} 0 & \text{if tail appears} \\ 1 & \text{if head appears} \end{cases}$$

then the random process is a family $\{X(1), X(2), X(3), X(4)\}$

and the State space $= \{0, 1\}$.

Markov chain: Consider a stochastic process associated with a parameter t, Let $t = 1, 2, 3 \ldots$ and let the corresponding random variables be $x_1, x_2, x_3, \ldots$. Let each random variable be discrete. Also let the state space of the process be $\{a_1, a_2, a_3 \ldots\}$.

Let the state space also be discrete let $X_{r+1} = a_j$ and $X_r = a_i$

Then the process for which $X_{r+1} = a_j$ occurs given that the process in which $X_r = a_i$ has already occured is called <u>the Markov chain</u>.

If the state space is a continious set then the corresponding markov chain is called markov process.

Transition Probabilities.

Let the probabilities corresponding to each pair of states (a_i, a_j) be P_{ij}

then $P_{ij} = P[X_{r+1} = a_j / X_r = a_i]$

The matrix $[P_{ij}]$ is called the transitional matrix or transition probability matrix [TPM]. Denoted by T.

Let the probabilities P_{ij} are called as Transitional probabilities.

Thus $P = [P_{ij}] = \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2n} \\ \vdots \\ P_{n1} & P_{n2} & \ldots & P_{nn} \end{bmatrix}$

Probability Vector:

A vector $V = (V_1, V_2, \ldots V_n)$ is called a probability vector if each of its components are non-negative and their sum $= 1$.

Stochastic Matrix: A square matrix $P = [P_{ij}]$ having every row in the form of probability vector is called stochastic Matrix.

Examples of Probability Vectors:

$V_1 = (0, 1)$ $V_2 = (1/3, 1/3, 1/3)$

Examples of stochastic matrix

$P_1 = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$ $P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Regular stochastic Matrix: A stochastic matrix P is said to regular if all the entries of P^n, $n = 1, 2, 3 \ldots$ are positive.

Ex: $P = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$

$P^2 = \frac{1}{2} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

$\therefore$ all the entries in P^2 are +ve $\therefore$ P is a regular stochastic matrix.

Irreducible markov chain: A markov chain is said to be Irreducible if its transitional probability matrix is a regular stochastic matrix.

Fixed Probability vector : Given a regular stochastic matrix P of order n if there exists a probability vector v of order $(1 \times n)$ such that $vP = v$ then v is called a fixed probability vector of the matrix P. It can be proved that v exists and it is unique.

PROBLEMS ON THE ABOVE

**

1. Verify that the matrix $P = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is a regular stochastic matrix.

sol^n : Given $P = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

Since all the entries in P are non-negative and sum of the entries of each row is 1, therfore P is a stochastic matrix.

$$P^2 = \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 3/8 & 5/8 & 0 \\ 1/4 & 5/8 & 1/8 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 5/16 & 19/32 & 3/32 \\ 5/16 & 5/8 & 1/16 \\ 1/4 & 5/8 & 1/8 \end{bmatrix}$$

2. Find the unique fixed probability vector of regular stochastic matrix $P = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}$

sol^n : $P = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}$

Let $v = (v_1, v_2)$ be the unique fixed probability vector such that $v_1 + v_2 = 1$ —— ① and $vP = v$

$$(v_1, v_2) \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix} = (v_1, v_2)$$

$$\frac{3}{4}v_1 + \frac{1}{2}v_2, \quad \frac{1}{4}v_1 + \frac{1}{2}v_2 = (v_1, v_2)$$

$$\frac{3}{4}v_1 + \frac{1}{2}v_2 = v_1$$

$$\left(\frac{3}{4} - 1\right)v_1 + \frac{1}{2}v_2 = 0$$

$$-\frac{1}{4}v_1 + \frac{1}{2}v_2 = 0 \quad \text{—— ②}$$

from ① and ②

$$\frac{1}{2}v_1 + \cancel{\frac{1}{2}v_2} = \frac{1}{2}$$

$$\underset{(+)}{-}\frac{1}{4}v_1 \underset{(-)}{+} \cancel{\frac{1}{2}v_2} = \underset{(-)}{0}$$

$$\frac{1}{2}v_1 + \frac{1}{4}v_1 = +\frac{1}{2}$$

$$\frac{6}{8}v_1 = \frac{1}{2}$$

$$v_1 = \frac{\cancel{8}\,2}{\cancel{12}\,3} \qquad v_2 = \frac{1}{3}$$

Is the unique fixed probability vector.

3). Find the unique fixed probability vector for regular stochastic matrix

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/6 & 1/2 & 1/3 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

Let $v = (v_1, v_2, v_3)$ be the unique fixed probability vector

$$v_1 + v_2 + v_3 = 1 \quad \text{—— ①}$$

and $vP = v$

$$[v_1 \ v_2 \ v_3]\begin{bmatrix} 0 & 1 & 0 \\ 1/6 & v_2 & 1/3 \\ 0 & 2/3 & v_3 \end{bmatrix} = (v_1, v_2, v_3)$$

$$\left[0v_1 + \frac{1}{6}v_2 + 0v_3, v_1 + \frac{1}{2}v_2 + \frac{2}{3}v_3, 0v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3\right] = [v_1, v_2, v_3]$$

$0v_1 + \frac{1}{6}v_2 + 0v_3 = v_1$ —— ②

$-v_1 + \frac{1}{6}v_2 + 0v_3 = 0$ —— ② and $v_1 + \frac{1}{2}v_2 + \frac{2}{3}v_3 = v_2$

$v_1 - \frac{1}{2}v_2 + \frac{2}{3}v_3 = 0$ —— ③

$v_1 = \frac{1}{10}$ $\quad v_2 = \frac{3}{5}$ $\quad v_3 = \frac{3}{10}$

from that ;

$V = (v_1, v_2, v_3) = \left[\frac{1}{10}, \frac{6}{10}, \frac{3}{10}\right]$ is the unique fixed probability vector.

④ Prove that the Markov chain whose TPM is $P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ is irreducible. Find the corresponding stationary probability vector.

Sol^n: Given : $P = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$

Since all the entries in P are non negative ie ≥ 0 and the sum of all the entries of each row is 1.

$\therefore$ P is a stochastic matrix

$$P^2 = \begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/6 & 1/3 \\ 1/4 & 7/12 & 1/6 \\ 1/4 & 1/3 & 5/12 \end{bmatrix}$$

Since all the entries in P are non-negative its regularly stochastic and hence markov chain is irreducible.

Let $V = v_1, v_2, v_3$ be the unique steady state probability vector such that $v_1 + v_2 + v_3 = 1$ —— ① and $VP = V$

$$(v_1, v_2, v_3)\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} = (v_1, v_2, v_3)$$

$$\left[0v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3, \frac{2}{3}v_1 + 0v_2, \frac{1}{2}v_3, \frac{1}{3}v_1 + \frac{1}{2}v_2 + 0v_3\right] = (v_1, v_2, v_3)$$

$0v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3 = v_1$; $\quad -v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3 = 0$ —— ②

$\frac{2}{3}v_1 + 0v_2 + \frac{1}{2}v_3 = v_2$; $\quad \frac{2}{3}v_1 - v_2 + \frac{1}{2}v_3 = 0$ —— ③

$\therefore$ from ①, ②, ③

$v_1 = \frac{1}{3}$ $\quad v_2 = \frac{10}{27}$ $\quad v_3 = \frac{8}{27}$

$V = (v_1, v_2, v_3) = \left[\frac{1}{3}, \frac{10}{27}, \frac{8}{27}\right]$ is the steady state probability vector.

⑤ Assume that a computer system is in one of the three states ; busy, idle, undergoing repair denoted by the states 0, 1, 2. Observing its state at a certain specified time at on each day it is found that the system approximately behaves like a Markov chain with the transition probability matrix $P = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.6 & 0 & 0.4 \end{bmatrix}$.

Prove that the chain is irreducible and determine the steady state probability vector.

Sol^n: $P = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.6 & 0 & 0.4 \end{bmatrix}$

Since all the entries are non-negative and the sum of the entries of each row is 1. $\therefore$ P is a stochastic matrix.

$$P^2 = \begin{bmatrix} 0.5 & 0.28 & 0.22 \\ 0.2 & 0.66 & 0.14 \\ 0.6 & 0.12 & 0.28 \end{bmatrix}$$

$\therefore$ P is a regular stochastic and hence markov chain is irreducible.

Let V be the unit steady state probability

$VP = V$

$$(v_1, v_2, v_3)\begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.6 & 0 & 0.4 \end{bmatrix} = (v_1, v_2, v_3)$$

$$\left[0.6v_1 + 0.1v_2 + 0.6v_3,\ 0.2v_1 + 0.8v_2 + 0v_3,\ 0.2v_1 + 0.1v_3 + 0.4v_3\right] = (v_1, v_2, v_3)$$

$0.6v_1 + 0.1v_2 + 0.6v_3 = v_1$; $-0.4v_1 + 0.1v_2 + 0.6v_3 = 0$ —— (2)

$0.2v_1 + 0.8v_2 + 0v_3 = v_2$; $0.2v_1 - 0.2v_2 + 0v_3 = 0$ —— (3)

from (1) (2) (3)

$v_1 = \frac{2}{5}$ $\quad v_1 = \frac{2}{5}$ $\quad v_3 = \frac{1}{5}$

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.6 & 0 & 0.4 \end{bmatrix} \end{array}$$?

**

(6) 3 boys A, B, C are throwing a ball to each other; A always throws the ball to B, B always throws the ball to C, but C is just as likely to throw the ball to B : A. If C was the first person to throw the ball find the probability that after 3 throws (i) A has the ball (ii) B has the ball (iii) C has the ball

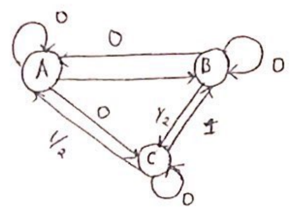
Soln: Let the state space of the system be $\{A, B, C\}$

where A: A is having the ball

B : B having the ball

$$P = \begin{array}{c} A \\ B \\ C \end{array}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 0 1/2 & 0 \end{bmatrix}$$

Transition diagram :



The initial Probability distribution is given by $A^0 = [p_1^{(0)}, p_2^{(0)}, p_3^{(0)}]$

$= [0, 0, 1]$

$$P^3 = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

$$\left[p_1^{(3)}, p_2^{(3)}, p_3^{(3)}\right] = A^{(0)} P^{(3)} = [0, 0, 1]\begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

$= [0.25, 0.25, 0.5]$

After 3 throws the Probability :

(i) 0.25. (ii) 0.25 (iii) 0.5.

(7) A software engineer goes to the work place everyday by motorbike or by car. He never goes by bike on two consecutive days but if he goes by car on a day then he is equally likely to go by car or bike on the next day. Find the TPM for the chain of the mode of transport he uses if car is used on the first day of the week. Find the probability that (i) bike is used (ii) Car is used on the 5^{th} day.

Soln: Let the state space of the system be $\{A, B\}$ where

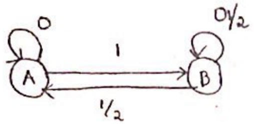
A : A by using bike

B : B by using car

the transitional probability Matrix is given by

$$P = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix} \end{array}$$

Transition diagram



$$P^4 = \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix}$$

Initial probability distribution is given by

$$A^{(0)} = [p_1^{(0)}, p_2^{(0)}] = [0, 1]$$

$$[p_1^{(4)}, p_2^{(4)}] = A^{(0)} P^{(4)}$$

$$= [0,1] \begin{bmatrix} 3/8 & 5/8 \\ 5/16 & 11/16 \end{bmatrix} = [5/16, 11/16]$$

On the fifth day (i) Probability bike is used is 5/16
(ii) Probability car is used is 11/16

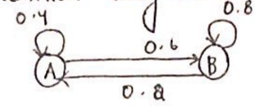
(8) A students study habits are as follows;
If he study on one night he is 60% ~~sure~~ sure not to study next night on the other hand if he does not study one night he is 80% sure not to study next night. In the long run how often does he study

Soln: Let the state space of the system be $\{A, B\}$

A : study B : no study

$$P = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \end{array}$$

Transition diagram



Let $V = (V_1, V_2)$ be the unique fixed probability vector

$V_1 + V_2 = 1$ —— (1) $VP = V$

$$[V_1, V_2] \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} = (V_1, V_2)$$

$$[0.4V_1 + 0.2V_2, \ 0.6V_1 + 0.8V_2] = [V_1, V_2]$$

$$0.4V_1 + 0.2V_2 = V_1$$

$$-0.6V_1 + 0.2V_2 = 0 \text{ —— (2)}$$

$V_1 = 1/4$ $V_2 = 3/4$

In the long run the student will study

$\frac{1}{4} \times 100 = 25\%$ of his time.

(9) A Company executive changes his car every year if he has a car of make A, he changes over to a car of make B. If he has a car of make B, ~~he changes over to a car~~ C, if he has a car of C he is just as likely to change over to a car of make C or B or A. If he has a car of make C in the year 2008. Find the probability that he will have a car of
(i) Make A in 2010 (ii) Make C in 2010 (iii) Make B in 2011
(iv) Make C in 2011
In the long run how often will he have a car of make C.

(i) 1/9 (ii) 4/9 (iii) 7/27
(iv) 16/27
[1/6, 1/3, 1/2]

(10) A travelling Salesman S sells in 3 cities A, B, C. He never sells in the same city on successive days. If he sells in city A on a day. Then the next day he sells in city B. However if he sells in either B or C then the next day he is twice as likely to sell in city A as in the other city. Find the TPM of the chain of selling. Find how often does the Salesman sells in city A, B and C.

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 2/3 & 0 & 1/3 \\ 2/3 & 1/3 & 0 \end{bmatrix}$$

soln: $[0.4, 0.45, 0.15]$

Scanned with CamScanner

Recurrent State : A state is said to be recurrent state if starting from the state A_i, the system returns to the same state A_i

Eg. Consider a transition probability matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

$a_1 = (0,1) \qquad a_2 = (1,0)$

$$P^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad P^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Since $P^3 = P$ therefore the system returns to same stage after 2 steps. Thus the state represented by the matrix P is recurrent state

Transient state : The state is said to be a transient state if the probability matrix a_j system and does not go back to a_j.

Example Let $P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0\,1/2 & 0 & 1/2 \end{bmatrix}$

$$P^2 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.5 & 0.25 & 0.25 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.25 & 0.375 & 0.375 \\ 0.375 & 0.25 & 0.375 \\ 0.375 & 0.375 & 0.25 \end{bmatrix}$$

We find that matrix P cannot be obtained through transitions, hence the states represented by matrix P are transient states

Absorbing states : The state a_i is said to be an A.S. if the transition probability is given by $P_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$

EX: $P = \begin{bmatrix} 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{bmatrix}$

The states a_2 and a_3 are absorbing states

CHI-SQUARE TEST $[\chi^2]$

When a fair coin is tossed 40 times we expect that head will appear 20 times and tail will appear 20 times. But in actual practice the experimental results do not agree with theoretical results. In all random Trials there exists some discrepancy between expected [theoretical and observed frequencies]

The discrepancy is analysed through a test statistic known as chi-square (χ^2) Let $O_1, O_2, O_3 \ldots\ldots O_N$ be the set of observed frequency and $E_1, E_2, E_3 \ldots\ldots E_N$ be the corresponding set of expected or theoretical frequencies then Chi-square is given by $\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

also $N = \sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i$

If the expected frequency are equal (atleast) to 5. Then it can be proved that sampling distribution of chi-square is approximately identical with probability distribution of chi-square whose density function is given by

$$P(\chi^2) = P_0 \chi^{2-2} e^{-\chi^2/2}$$

where ? is a positive constant and P_0 is a constant depending on ? such that the total area under the curve is 1.

If calculated value of chi-square is less than the tabulated value of χ^2 then the value of χ^2 is not significant and the hypothesis of can be accepted.

If the calculated value of chi-square is greater than the tabulated value of χ^2, then we conclude that the observed frequency differ significantly from the expected frequency and would reject the hypothesis at a corresponding level of C.

Degree of freedom: The no. of degree of freedom is given by $\nu = n - m$ where n is the no. of frequency pairs (O_i, E_i) and m is the no. of quantities needed for (used for) calculating E_i.

If $N = \sum O_i$ is the only quantity used in the calculation of E_i then $m = 1$ and $\nu = n - 1$

Goodness of fit: When a hypothesis is accepted on the basis of chi-square test we say that the expected frequency calculated on the basis of hypothesis form a good fit for the given frequencies. When the hypothesis is rejected we say that the corresponding expected frequencies do not form a good fit.

**
Problems on Chi-Square Test

① In an experiment on pea breeding the following frequencies of seeds were obtained

Round and Yellow – 315.0 , Wrinkled and Yellow – 101

Round and Green – 108 , Wrinkled and Green – 32. Total – 556

Theory predicts that the frequency should be in proportion 9:3:3:1. Examine the correspondance between theory and experiment.

Sol^n: $O_1 = 315$ $O_2 = 101$, $O_3 = 108$ $O_4 = 32$

$$N = \sum_{i=1}^{4} O_i = O_1 + O_2 + O_3 + O_4 = 315 + 101 + 108 + 32 = 556$$

$$E_1 = \frac{9}{16} \times N = \frac{9}{16} \times 556 = 313$$

$$E_2 = \frac{3}{16} \times N = \frac{3}{16} \times 556 = 104$$

$$E_3 = \frac{3}{16} \times N = \frac{3}{16} \times 556 = 104$$

$$E_4 = \frac{3}{16} \times N = \frac{1}{16} \times 556 = 35$$

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$= \frac{4}{313} + \frac{9}{104} + \frac{16}{104} + \frac{9}{35}$$

$$= 0.0127 + 0.0865 + 0.1538 + 0.2571.$$

$$\chi^2 = 0.5101.$$

$\nu = n - 1 = 4 - 1 = 3.$ $\chi^2_{0.05}(3) = 7.815$

Since the calculated value of chi-square is less than the tabulated value of χ^2. Therefore value is non-significant and the hypothesis that the theory agrees with the experiment can be accepted at the 0.05 degree of significance.

② Genetic theory states that children having one parent of blood type M and other parent of blood type N will always have one of the three types of blood.

(i) M (ii) MN (iii) N and that the proportion of these types will be 1:2:1. A report states that out of 300 children having one M parent and one N parent 30% were found to be of type M and 45% of type MN and the remainder of type N. Test the hypothesis by chi-square test

Sol^n:

	M	MN	N
O_i	30% of 300 = 90	45% of 300 = 135	25% of 300 = 75.
E_i	$\frac{1}{4} \times 300$ = 75	$\frac{2}{4} \times 300$ = 150	$\frac{1}{4} \times 300$ = 75
$O_i - E_i$	15	-15	0
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{225}{75}$	$\frac{225}{135}$	$\frac{\cancel{0}}{75}$ 0

$\chi^2_{0.05}(2) = 5.99$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{225}{75} + \frac{225}{135} + 0 = 4.6$$

$\gamma = n-1 = 3-1 = 2.$

Since the calculated value of chi-square is less than the tabulated value of chi-square therefore χ^2 is not significant and the hypothesis that agrees with the theory can be expected at 0.05 significance

③ The following table gives the no. of aircraft accidents that occured during various days of the week. Find whether the accidents are uniformly distributed over week.

Days :	Sunday	Monday	Tuesday	Wednesday	thursday	friday	saturday
No. of Accidents :	14	16	8	12	11	9	14

Total no. of accidents = 84.

O_i =	14	16	8	12	11	9	14
$E_i = \frac{1}{7} \times 84 = 12$	12	12	12	12	12	12	
$O_i - E_i$	2	4	-4	0	-1	-3	-2
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{4}{12}$	$\frac{16}{12}$	$\frac{16}{12}$	$\frac{0}{12}$	$\frac{1}{12}$	$\frac{9}{12}$	$\frac{4}{12}$

$$\chi^2 = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} = 4.16$$

$\gamma = n-1 = 7-1 = 6$

$\chi^2_{0.05}(6) = 12.59$

$\chi^2_{0.01}(6) = 16.81$

Since the calculated value of χ^2 is less than the tabulated value of χ^2. Therefore the value of χ^2 is not significant and we do not reject the hypothesis that the accidents are seemed to be uniformly distributed over the γ at 0.05 and 0.01 level of significance

④ A die was thrown 60 times and the following frequency distribution is considered. Faces 1, 2, 3, 4, 5, 6, frequency 15, 6, 4, 7, 11, 17. Test whether the dice is unbiased.

sol^n:

faces	1	2	3	4	5	6	Total
O_i	15	6	4	7	11	17	60
$E_i = \frac{1}{6} \times 60 = 10$	10	10	10	10	10	10	60
$O_i - E_i$	25	-4	-6	-3	1	7	
$\frac{(O_i - E_i)^2}{E_i}$	2.5	1.6	3.6	0.9	0.1	4.9	

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 13.6$$

$\gamma = n-1 = 6-1 = 5$

$\chi^2_{0.05}(5) = 11.07$ $\qquad$ $\chi^2_{(0.01)}(5) = 15.09$

Since the calculated value of χ^2 is greater than the tabulated value. Therefore the value of χ^2 is significant and the hypothesis that the die is unbiased χ^2 is rejected

Since the value of χ^2 is less than the tabulated value at 0.01 level of significance therefore the value of χ^2 is not significant and the hypothesis that die is unbiased is accepted

⑤ A set of five identical coin are tossed 320 times and the result is as follows

1. No of heads	0	1	2	3	4	5
2. Frequency	6	27	72	112	71	32

Test the hypothesis that the data follows a binomial distribution given $\chi^2(0.05)(5) = 11.07$ $\quad \chi^2_{0.01}(5) = 15.09$

sol^n:

No of heads	0	1	2	3	4	5	Total
Frequency (O_i)	6	27	72	112	71	32	320
$E_i = N\,b(x)$	10	50	100	100	50	10	320

$$= \left[320\ {}^5C_x \left(\tfrac{1}{2}\right)^x \left(\tfrac{1}{2}\right)^{5-x}\right]$$

$O_i - E_i$ = -4 -23 -28 12 21 22

③ $\frac{(O_i - E_i)^2}{E_i}$ = $\frac{16}{10}$ $\frac{529}{50}$ $\frac{784}{100}$ $\frac{144}{100}$ $\frac{441}{50}$ $\frac{484}{10}$

$\nu = n-1 = 6-1 = 5$

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 78.68$$

5 $\chi^2_{0.05}(5) = 11.07$ $\chi^2_{0.01}(5) = 15.09$

Since the calculated value of χ^2 is greater than the tabulated value of χ^2 at 0.05 and 0.01 level of significance therefore the value of χ^2 is significant and the hypothesis that the data follows binomial distribution associated with fair coin cannot be accepted

⑥ A survey of 320 families with 5 children each has the following distribution.

No. of boys/Girl	5	4	3	2	1	0
No. of families	14	56	110	88	40	12

Is this data consistant with the hypothesis that the male and female births are equally probable. Use χ^2 test at 0.05 and 0.01 level of significance [$\chi^2_{0.05}(5) = 11.07$, $\chi^2_{0.01}(5) = 15.09$]

sol^n:

No of girls	0	1	2	3	4	5	Total
No of families (O_i)	14	56	110	88	40	12	320
$E_i = N\,b(x)$ $= 320\ {}^5C_x (0.5)^x (0.5)^{5-x}$	10	50	100	100	50	10	320
$O_i - E_i$	4	6	10	-12	-10	2	
$\frac{(O_i - E_i)^2}{E_i}$	$\frac{16}{10}$	$\frac{36}{50}$	$\frac{100}{100}$	$\frac{144}{100}$	$\frac{100}{50}$	$\frac{4}{10}$	7.16

$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 7.16$$

$\nu = n-1 = 6-1 = 5$ $\chi^2_{0.05}(5) = 11.07$ $\chi^2_{0.01}(5) = 15.09$

The value of χ^2 is not significant and the data is consistant with the hypothesis that male and female births are equally probable

STUDENT'S - t TEST

Consider a small sample of size s ($N \leq 30$); drawn from a normal population with mean (μ) and standard deviation σ. If $\bar{x}$ and s are the sample mean and the sample standard deviation then the statistical value t is defined as

$$t = \frac{(\bar{x} - \mu)}{s} \sqrt{N-1}$$

$\bar{x}$ and s are the statistics and $\mu = N-1$ is the degree of freedom of t. If we calculate 't' for each sample the value. value we obtain in sampling distribution for t is known as students t distribution. The curve is represented by the formula

$$Y(t) = Y_0 \left[1 + \frac{t^2}{N-1} \right]^{-N/2}$$

where Y_0 is an appropriate constant called at 't' term and Y_0 is so choosen that total area under the curve is 1.

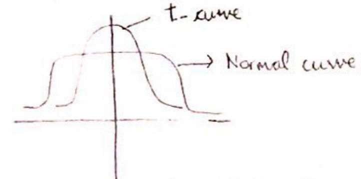
For large values of N, Y_t reduces to $\phi(z)$

Significant test of sample.

Let $x_1, x_2, \ldots \ldots x_n$ be the random small samples from the normal population. To test the normal population the hypothesis the mean (μ) and t is calculated

$$t = \frac{(\bar{x} - \mu)}{s} \sqrt{N-1} \qquad \bar{x} = \sum \frac{x_i}{N}$$

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{N}$$



If the calculated value of t is greater than the tabulated value, at a particular level of significance then the difference between $\bar{\mu}$ and μ at that level of significance. If the calculated value of t is less than the tabulated value of t at a particular level if with the hypothesis that mean of population is μ.

(2)

If two sets of data are given with the same sample size then t is calculated by using

$$t = \frac{(\bar{d} - \mu)}{s} \sqrt{N-1}$$

where $d = y - x$ $\quad \bar{d} = \sum \frac{d_i}{N}$

$$s^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{N}$$

If two sets of data are given with different sample size then t is calculated using

$$t = \frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

$$\bar{x} = \sum \frac{x_i}{N_1} \qquad \bar{y} = \sum \frac{y_i}{N_2}$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{(N_1 + N_2 - 2)}$$

PROBLEMS ON STUDENTS - t Test

(1) A certain stimulus administered to each of 12 patients resulted in change in the following blood pressure. The readings are as such

5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4, 6

Can it be concluded that the stimulus will in general be accompanied by a change in the blood pressure. Given $t_{0.05}(11) = 2.2$

Soln $\bar{x} = \sum \frac{x_i}{N} = \frac{5+2+8-1+3+0-2+1+5+0+4+6}{12}$

$$\bar{x} = \frac{31}{12} = 2.58$$

$$s^2 = \sum \frac{(x_i - \bar{x})^2}{N} = \frac{(5-2.58)^2 + (2-2.58)^2 + (8-2.58)^2 + (-1-2.58)^2 + (3-2.58)^2 + (-2-2.58)^2 + (1-2.58)^2 + (5-2.58)^2 + (0-2.58)^2 + (0-2.58)^2 + (4-2.58)^2 + (6-2.58)^2}{12}$$

$s^2 = 8.74 \Rightarrow s = 2.96$

$$t = \frac{\bar{x} - \mu}{s} \sqrt{N-1} = \frac{2.58 - 0}{2.96} \sqrt{12-1}$$

$t = 2.89$

$t_{0.05}(11) = 2.2$ (Given)

Therefore the value is greater than the tabulated value. Therefore it cannot be accepted at 5% level of significance.

(2) The 9 items of the sample have the following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these differs significantly from assumed mean of 47.5. Given $t_{0.05}(8) = 2.31$.

Soln $\bar{x} = \sum \frac{x_i}{N} = \frac{45+47+50+52+48+47+49+53+51}{9}$

$\bar{x} = 49.11$

$$s^2 = \sum \frac{(x_i - \bar{x})^2}{N} = \frac{(45-49.11)^2 + (47-49.11)^2 + (50-49.11)^2 + (52-49.11)^2 + (48-49.11)^2 + (47-49.11)^2 + (49-49.11)^2 + (53-49.11)^2 + (51-49.11)^2}{9}$$

$s^2 = 6.018 \Rightarrow s = 2.46$

$$t = \frac{(\bar{x} - \mu)}{s} \sqrt{N-1} = \frac{49.11 - 47.5}{2.46} \sqrt{8}$$

$= 1.85$

$t_{0.05}(8) = 2.31$

Since the calculated value of t is less than tabulated value of t at 0.05 significant value

Therefore the hypothesis is accepted, i.e. there is no significant difference between the sample mean and population mean at 0.05 level of significance

(3) A mechanist is making engine parts with axial dia of 0.7 inch. A random sample of 10 parts shows mean diameter of 0.742 inch with a deviation of 0.04 inch. On the basis of this sample would you say that the work is inferior. Given that $t_{0.059} = 2.26$

Soln: $\bar{x} = 0.742$ $\quad s = 0.04$

$\mu = 0.7$ $\quad N = 10$

$$t = \frac{\bar{x} - \mu}{s}\sqrt{N-1} = \frac{(0.742 - 0.7)\sqrt{9}}{0.04} = 3.15$$

$$t = 3.15$$

$$t_{0.059} = 2.26$$

Since the calculated value of t is greater than the tabulated value of t therefore the value of t is significant and the hypothesis is rejected.

Thus we conclude that at 5% level of significance, the work is inferior.

— EXTRA PAGE —

$t = \bar{y} - \bar{x}$ 15 - 13 ... were given a month

11 - students were given a test in stanes. ... were given a month further tution and a second test of equal difficulty was held at the end of it. Do the marks give evidence that the students have been benifited by extra tutorial.

Boys :	1	2	3	4	5	6	7	8	9	10	11
Marks 1 :	23	20	19	21	18	20	18	17	23	16	19
Marks 2 :	24	19	22	18	20	22	20	20	23	20	17

Solⁿ : $\bar{d} = \frac{\Sigma(y-x)}{N} = \frac{(24-23) + (19-20) + (22-19) + (18-21) + (20-18) + (22-20) + (20-18) + (20-17) + (23-23) + (20-16) + (17-19)}{11}$

$\bar{d} = \frac{1-1+3-3+2+2+2+3+0+4-2}{11} = \frac{11}{11} = 1$

$\bar{d} = 1$

$s^2 = \frac{\Sigma(d_i - \bar{d})^2}{N} = 4.545 \quad \therefore \quad s = \sqrt{4.545} = 2.13$

$t = \frac{(\bar{d} - \mu)}{s}\sqrt{N-1}$

$= \frac{1}{2.13}\sqrt{11-1} = \frac{\sqrt{10}}{2.13} = 1.485$

$\nu = n-1 = 11-1 = 10$

$t_{0.05}(10) = 2.228$

Since the calculated value of t is less than the tabulated value of t, therefore it is not significant do not reject the hypothesis that the student get extra benefit from the coaching at 5% significance.

Objectives of Sampling: From a random sample of 10 cows fed on diet "A" the increase in weight in a certain period is 10, 6, 17, 16, 13, 12, 8, 14, 15, 9. From another sample of 12 cows for diet B the increase in weight in same period is 7, 13, 22, 15, 12, 14, 18, 8, 21, 23, 10, 17. Test whether diets A and B differ significantly with regard to their effect on weights.

Solⁿ : x_i 10 6 17 16 13 12 8 14 15 9

y_i 7 13 22 15 12 14 18 8 21 23 10 17

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
10	7	-2	4	-8	64
6	13	-6	36	-2	4
17	22	4	16	7	49
16	15	5	25	0	0
13	12	1	1	-3	9
12	14	0	0	-1	1
8	18	-4	16	3	9
14	8	2	4	-7	49
15	21	3	9	6	36
9	23	-3	9	8	64
	10			-5	25
	17			2	4
$\Sigma x_i = 120$	$\Sigma y_i = 180$		$\Sigma(x_i - \bar{x})^2 = 120$		$\Sigma(y_i - \bar{y})^2 = 314$

$\bar{x} = \frac{\Sigma x_i}{N_1} = \frac{120}{10} = 12 \qquad \bar{y} = \frac{\Sigma y_i}{N_2} = \frac{180}{12} = 15$

$\sigma^2 = \frac{\Sigma\left[(x_i - \bar{x})^2\right] + \Sigma(y_i - \bar{y})^2}{(N_1 + N_2 - 2)} = \frac{120 + 314}{10 + 12 - 2} = \frac{434}{20} = 21.7$

$\sigma = 4.66$.

Scanned with CamScanner

$$t = \frac{\bar{y} - \bar{x}}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{15 - 12}{\sigma\sqrt{\frac{1}{10} + \frac{1}{12}}} = \frac{3}{\sigma\sqrt{0.1 + 0.083}}$$

$$= \frac{3}{4.66\sqrt{0.1 + 0.083}} = 1.5049$$

$$\nu = N_1 + N_2 - 2 = 10 + 12 - 2 = 20$$

$$t_{0.05}(20) = 2.78$$

Since the calculated value of t is less than the tabulated value, therefore the value of t is not significant, hence their is no significant difference with regard to effect of A and B diets.

SAMPLE AND SAMPLING

A small section selected from a large group (population) is called sample and the process of drawing a sample is called sampling.
Population can be finite or infinite.
The process of drawing a sample such that each member of population has equal chance of being included in the sample is called random sampling.

Simple sampling is a special case of random sampling in which each event has the same probability of success 'p' and the chance of success of event are independent whether previous trials have been made or not.

Statistical Constants of the population such as mean and standard deviation are called the parameters of the population, whereas the statistical constants for the sample drawn from a given population mean, SD, etc are called statistics.
The population parameters are not generally known and their estimates given by the statistical analysis are used.

- At 0.05 level of significance, $z_c = 1.96$
  $z = 1.6 \in (-1.96, 1.96)$

Objectives of Sampling : Sampling aims at gathering maximum info. with minimum effort, cost and time. The objective of sampling is to obtain the best possible values of the parameters under specific conditions. The logic of sampling theory is the logic of induction in which we pass from a particular sample to a general population. Such a generalization from sample to population is called statistical infrences.

Sampling distributions :
Consider all possible samples of size N which can be drawn from a given population at random. For each sample compute the mean. If we group these means according to frequencies, then the frequency distribution is known as sampling distribution of means.
Illy we can have sampling distribution of standard deviations.
While drawing each sample, we put back the previous sample, such that the parent population remains the same, then such sampling is called sampling with replacement.

Standard Error : The standard deviation of sampling distribution is called standard Error.
The standard error of sampling distribution of means is called standard error of means.
The standard error is used to assess the difference between the observed and expected values.
The reciprocal of standard error is called precision.
If $N \geq 30$ the sample is called a large sample otherwise small.
The sampling distribution of large samples is assumed to be normal.

Testing of Hypothesis
To reach decision about population about on the basis of sample information, we make assumption about the population involved. Such assumption which may or may not be true is called as statistical hypothesis.

* The statistical hypothesis formulated for the sake of rejecting it under the assumption that it is true it is called Null hypothesis.

Alternate hypothesis :
The statistical hypothesis which is different from Null hypothesis is called alternate hypothesis.

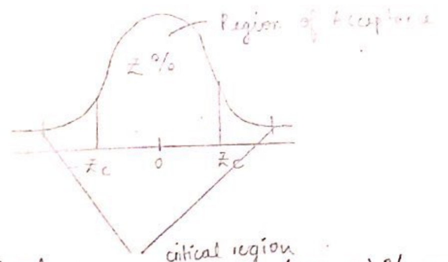
TYPE I AND TYPE II ERRORS :

	ACCEPTED	REJECTED
H_0 is true	No error	Type I error
H_0 is false	Type II Error	No error

The statistical testing of hypothesis aims at minimizing type I error to a pre-assigned value 5% or 1% and to minimize type II error.
The only way to reduce both types of error is to increase the sample size.

** Level of significance
The probability below which we reject the hypothesis is called the level of significance.

** Critical region : The set of values of z that are outside the interval $(-z_c, z_c)$ is called the critical region or the region of rejection or the region of significance.



z % confidence level corresponds to $(100-z)\%$ or $1-z\%$ level of significance of normal curve.

level of significance	Critical value / confidence coefficient	Confidence interval
0.05	1.96	−1.96, 1.96
0.01	2.58	−2.58, 2.58

$z = \frac{S-\mu_s}{\sigma_s}$ is called the z-score of the statistic (s). It is this score, that determines the fate of a hypothesis and is called test statistic.

TWO TAILED TEST (Rule of decision)
Reject a hypothesis h at a α % is called the β level of hypothesis if z-score of the that is in a critical region $(-z_c, z_c)$.
Do not reject the hypothesis if inside.

Test of significance
The procedure which enables us to decide whether to accept or reject the hypothesis is called the test of significance or test of hypothesis or rule of decision.

Sampling distribution of means
$\mu_{\bar{x}} = \mu$

(i) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \frac{\sqrt{N_p - N}}{\sqrt{N_p - 1}}$

N_p – population size
N – sample size
if population in finite

(2) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ if the population is infinite

$z = (\bar{x} - \mu_{\bar{x}}) / \sigma_{\bar{x}}$ $(\bar{x} - \mu_{\bar{x}})$

Sampling distribution of proportions.

$\mu_p = p$

$\sigma_p = \sqrt{\frac{pq}{N}}$

$z = \frac{P - \mu_p}{\sigma_p}$

Two- Tailed test

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$Z = \frac{X - \mu_x}{\sigma_x}$$

Where $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$

$\mu_x = Np$ $\sigma_x = \sqrt{Np(1-p)}$

Interval estimates , confidence limits , confidence levels.

The procedure to determine the interval in which population parameter lie is called interval estimation and the interval is called confidence interval for that population interval and the end point of interval is called confidence limit.

Confidence limit for Mean is given by

$$\bar{X} \pm z_c \sigma_{\bar{x}} = \bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

Confidence limit for proportion is given by

$$p \pm z_c \sigma_p = p \pm z_c \sqrt{\frac{pq}{N}} = p \pm z_c \sqrt{\frac{p(1-p)}{N}}$$

A coin was tossed 400 times , heads turned up 216 times. Test the hypothesis that the coin is unbiased at 0.05 level of significance.

H_0 is such that coin is unbiased.

soln:

$X = 216$ $p = \frac{1}{2}$ $q = 1-p = \frac{1}{2}$ $N = 400$

$\mu_x = Np = 400 \times \frac{1}{2} = 200$

$\sigma_x = \sqrt{Npq} = \sqrt{400 \times \frac{1}{2} \times \frac{1}{2}} = 10$

$Z = \frac{X - \mu_x}{\sigma_x} = \frac{216 - 200}{10} = \frac{16}{10} = 1.6$

At 0.05 level of significance , $z_c = 1.96$

$Z = 1.6 \in (-1.96, 1.96)$

$\therefore$ The null hypothesis that the coin is unbiased can be accepted at 0.05 level

② A die was thrown 9000 , and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throw do the data indicated a unbaised one.

soln: $N = 9000$ $p = \frac{2}{6} = \frac{1}{3}$ $q = 1 - \frac{1}{3} = \frac{2}{3}$ $X = 3240$

$\mu_x = Np = 9000 \times \frac{1}{3} = 3000$

$\sigma_x = \sqrt{Npq} = \sqrt{\frac{9000}{1000} \times \frac{1}{3} \times \frac{2}{3}} = \sqrt{2000} = 20\sqrt{5} = 44.7213$

$Z = \frac{X - \mu_x}{\sigma_x} = \frac{3240 - 3000}{44.7213} = 5.366$

At 0.05 level of significance $z_c = 1.96$

At 0.01 level of significance $z_c = 2.58$

$Z = (-1.96, 1.96)$ and $(-2.58, 2.58)$.

Therefore the hypothesis is rejected at 0.05 and 0.01 by two tail test.

③ The mean life of a sample of 100 fluorescent tubelights manufactured by a company is found to be 175 hrs with a standard deviation of 120 hrs Test the hypothesis that the mean life time of the lights is 1600 hrs.

soln: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ $H_0 = \mu_{\bar{x}} = 1600$

$\mu_{\bar{x}} = \mu = 1600$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = \frac{120}{\sqrt{100}} = 12$

$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{1570 - 1600}{12} = -2.5$

0.05 level of significance $z_c = 1.96$

0.01 level of significance $z_c = 2.58$

$Z = -2.5 \notin (-1.96, 1.96)$ $Z = 2.5 \in (-2.58, 2.58)$

∴ the mean lifetimes of the light of 1600 cannot be accepted at 0.05 level of significance, whereas it can be accepted at 0.01 level of significance

④ A sample of 900 items is found of the mean equal to 3.4. Can it be reasonably regarded as a truely random sample from a large population with mean 3.25 and standard deviation 1.61.

Solⁿ: $\bar{x} = 3.4$ $\quad \mu_{\bar{x}} = 3.25$ $\quad \sigma = 1.61$ $\quad N = 900$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = ?$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} = \frac{1.61}{\sqrt{900}} = 0.054$$

$$z = \frac{3.4 - 3.25}{0.054} = 2.77$$

At 0.05 level and at 0.01 level of significance it is rejected as $z = 2.77$

⑤ Find how many heads in 64 tosses of a coin will ensure its fairness at 0.05 level of significance.

Solⁿ: $N = 64$ $\quad p = \frac{1}{2}$ $\quad q = \frac{1}{2}$

$$z = \frac{x - Np}{\sqrt{Npq}} = \frac{x - 64 \times \frac{1}{2}}{\sqrt{64 \times \frac{1}{2} \times \frac{1}{2}}} = \frac{x - 32}{4}$$

At 0.05 level of significance $z_c = 1.96$

$$-1.96 < z < 1.96$$

$$-1.96 < \frac{x - 32}{4} < 1.96$$

$$-7.84 < x - 32 < 7.84$$

$$32 - 7.84 < x < 7.84 + 32$$

$$25 < x < 39$$