PROBABILITY THEORY III

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

Continuous Probability Distribution & Bivariate Probability Distn

Some Continuous Theoritical Distributions discussed here:

i) Uniform Distribution ('Rectangular Distribution).
ii) Gamma Distribution.
iii) Beta Distribution.
iv) Exponential Distribution.
v) Normal Distribution.
vi) Double exponential or Laplace Distribution.
vii) Truncated normal distribution.
viii) Log-normal distribution.
ix) Pareto Distribution.
x) Cauchy Distribution.
xi) Logistic Distribution.

Scanned by CamScanner

SOME CONTINUOUS DISTRIBUTIONS

Rectangular Distribution or UNIFORM DISTRIBUTION :—

An absolutely continuous random variable X defined over $[a,b]$, $-\infty < a < b < \infty$ is said to follow uniform distribution with parameter a, b; if its pdf is given by,

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b, \\ 0, & \text{ow} \end{cases}$$

We will write $X \sim U[a,b]$ if X has a uniform distribution on $[a,b]$

This distribution is also called a rectangular distribution since the area under 'f' in between a and b is rectangular. It is also called Rectangular Distribution.

$$X \sim U[a,b]$$
or,
$$X \sim R[a,b]$$

The end point a or b or both may be excluded. Clearly,

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

Distribution Function :— The DF of X is given by,—

$$F(x) = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \ge b \end{cases};$$

Expectation & Variance :—

$$E(X^k) = \int_a^b x^k f(x)dx = \frac{1}{b-a}\int_a^b x^k dx \quad , k > 0 \text{ is an integer.}$$

$$= \frac{1}{b-a}\left[x^{k+1}\right]_a^b / k+1$$

$$= \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}.$$

Putting $k=1$, $E(X) = \frac{b+a}{2}$,

Putting $k=2$, $E(X^2) = \frac{(b^2+ab+a^2)}{3}$

$$\therefore Var(X) = E(X^2) - E^2(X) = \frac{(b-a)^2}{12}.$$

☒ Moment Generating Function: [C.U. 2003]

$$M_X(t) = E(e^{tX}) = \int_a^b e^{tx} dx \cdot \frac{1}{b-a}$$

$$= \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$$= \frac{1}{t(b-a)} \left[\sum_{j=0}^{\infty} \frac{(tb)^j}{j!} - \sum_{j=0}^{\infty} \frac{(ta)^j}{j!} \right]$$

$$= \frac{1}{t(b-a)} \sum_{j=0}^{\infty} \frac{(tb)^j - (ta)^j}{j!}$$

$$= \frac{1}{t(b-a)} \sum_{j=0}^{\infty} \frac{t^j}{j!} (b^j - a^j)$$

$$\therefore \mu_1' = E(X) = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}.$$

$$\therefore \mu_2' = E(X^2) = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

$$\therefore V(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

$$S.d. = \frac{|b-a|}{\sqrt{12}} = \frac{b-a}{\sqrt{12}}$$

NOTE: → The distribution is trivially symmetric about $\frac{a+b}{2}$.

⇨ Theorem: 1. (Probability Integral Transformation)

Let X be a continuous R.V. having D.F. F. Then $F(X)$ has the uniform distribution on $[0,1]$.

Proof:– Let, $Y = F(X)$

Then the DF of Y is given by —

$$F_Y(y) = P[Y \le y]$$

$$= P[F(X) \le y]$$

$$= P[X \le F^{-1}(y)]$$

$$= F[F^{-1}(y)] \quad [\text{as } F(x) \text{ is a monotone non-decreas function}]$$

$$= \begin{cases} 0 & \text{if } y \le 0 \\ 1 & \text{if } 0 < y < 1 \\ 1 & \text{if } y \ge 1 \end{cases}$$

$$\text{Pdf. of } Y = \frac{d}{dy}[G(y)] = 1$$

$$= \frac{d}{dy} \int_0^y g(y)\, dy$$

$$\therefore Y = F(X) \sim U(0,1).$$

Note that, $E[Y] = \frac{1}{2}$,
$V[Y] = \frac{1}{12}$.

NOTE:- The fact can be used to draw random observations from the theoritical distribution of X.

Here at first we choose 3 digited random numbers and put a decimal point before the first digit. Let us denote such a quantity by p, clearly p is a realization from $R(0,1)$ distn., now to obtain x we equate $F(x)=p$ and solve for x.

Theorem: 2. Let F be any DF, and let X be a $U[0,1]$ RV. Then there exists a function h such that $h(X)$ has DF F, i.e.,

$$P\{h(X) \le x\} = F(x) \quad \text{for all } x \in (-\infty, \infty).$$

Proof:- If F is the DF of a discrete RV Y, let

$$P[Y = y_k] = p_k, \quad k=1,2,\dots$$

Define h as follows:-

$$h(x) = \begin{cases} y_1 & \text{if } 0 \le x < p_1, \\ y_2 & \text{if } p_1 \le x < p_1+p_2, \\ \vdots & \vdots \end{cases}$$

Then

$$P\{h(X) = y_1\} = P\{0 \le X < p_1\} = p_1,$$
$$P\{h(X) = y_2\} = P\{p_1 \le X < p_1+p_2\} = p_2,$$

and in general,

$$P\{h(X) = y_k\} = p_k, \quad k=1,2,\dots$$

Thus $h(X)$ is a discrete RV with DF F.

If F is continuous and strictly increasing, F^{-1} is well defined, and we take $h(X) = F^{-1}(X)$. We have

$$P\{h(X) \le x\} = P\{F^{-1}(X) \le x\}$$
$$= P\{X \le F(x)\}$$
$$= F(x),$$

as asserted.

In general, define

$$F^{-1}(y) = \inf\{x : F(x) \ge y\},$$

and let $h(X) = F^{-1}(X)$. Then we have

$$\{F^{-1}(y) \le x\} = \{y \le F(x)\}.$$

$F^{-1}(y) \le x \Rightarrow \forall\ \epsilon > 0$, $y \le F(x+\epsilon)$, since $\epsilon > 0$ is arbitrary and F is continuous on the right, we let $\epsilon \to 0$ and conclude that $y \le F(x)$.

Since $y \le F(x) \Rightarrow F^{-1}(y) \le x$. Thus,

$$P[F^{-1}(X) \le x] = P[X \le F(x)] = F(x).$$

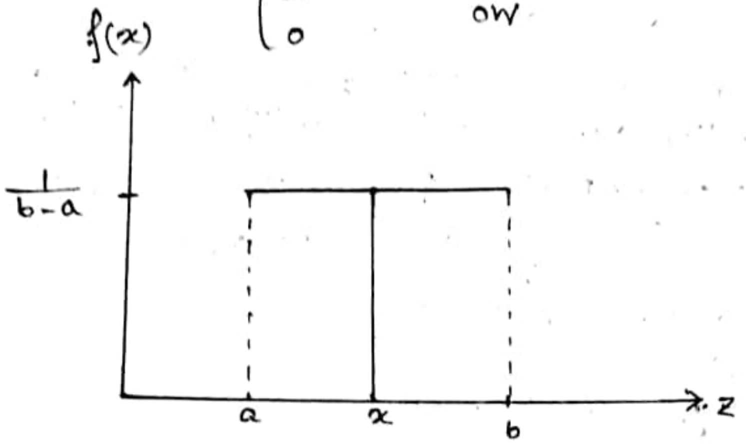
NOTE:- It is quite useful theorem in generating samples with the help of the uniform distribution.

Scanned by CamScanner

☆ **Problem:** → Why is the distribution called rectangular distn.?

Ans:-

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



The graph of the pdf $f(x)$ looks like a rectangular and the uniform distribution over $[a,b]$ is also known as Rectangular Distribution.

$$X \sim U[a,b]$$
or
$$X \sim R[a,b]$$

Remark:-

Let $X \sim U[\frac{1}{2}, 1]$

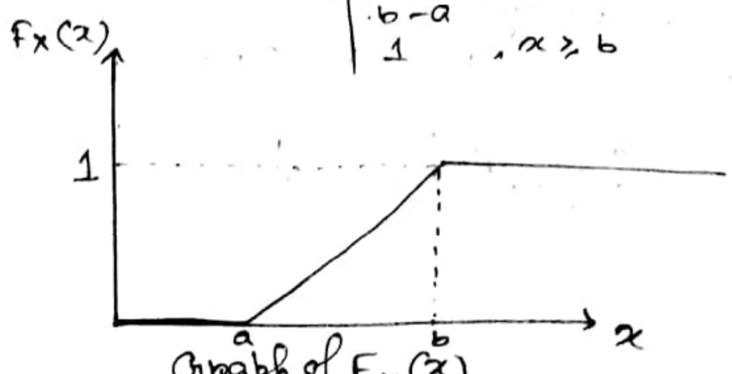
Then the p.d.f. of X is $f(x) = \begin{cases} 2 & , \frac{1}{2} < x < 1 \\ 0 & , \text{ow} \end{cases}$

Note that the p.d.f. $f(x)$ takes values greater than unity. In case of Uniform distribution,

$$\frac{1}{b-a} = \frac{1}{\text{length of the interval}}$$

CDF:- $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & , x \le a \\ \int_a^x \frac{1}{b-a}\,dt & , a < x < b \\ 1 & , x \geq b \end{cases}$

$$= \begin{cases} 0 & , x \le a \\ \frac{x-a}{b-a} & , a < x < b \\ 1 & , x \geq b \end{cases}$$



Graph of $F_X(x)$

GAMMA DISTRIBUTION :—

- **Definition:**– An absolutely continuous random variable X defined over $[0,\infty)$ is said to follow Gamma distribution with parameters α and p if its pdf is given by,—

$$f(x) = \frac{\alpha^p . e^{-\alpha x}}{\Gamma(p)} . x^{p-1} \quad \text{for } x>0 \;;\; \alpha, p>0$$
$$= 0 \qquad \text{OW}$$

where, $\Gamma(p)$ is of the form, $\Gamma(p) = \int_0^\infty e^{-t} t^{p-1} dt$ is the Gamma function.

We write, $X \sim G(\alpha, p)$.

Derivation of the p.d.f. :—
A probability law can easily be obtained from the following improper integral: $\int_0^\infty e^{-\alpha x} x^{p-1} dx$

Here the improper integral converges iff $\alpha>0 ; p>0$.

Now,
$$\int_0^\infty e^{-\alpha x} x^{p-1} dx = \frac{1}{\alpha}\int_0^\infty e^{-z}\left(\frac{z}{\alpha}\right)^{p-1} dz \qquad [\text{let, } \alpha x = z]$$
$$= \frac{1}{\alpha^p}\int_0^\infty e^{-z} z^{p-1} dz$$
$$= \frac{\Gamma(p)}{\alpha^p}$$

Now, $\int_0^\infty \frac{\alpha^p}{\Gamma(p)}\left(e^{-\alpha x} . x^{p-1}\right) dx = 1 \Rightarrow$ It is verified that it's a p.d.f.

We can define, $f(x) = \frac{\alpha^p . e^{-\alpha x} . x^{p-1}}{\Gamma(p)} I_x(0,\infty)$, where $\alpha>0, x>0, p>0$.

Since the p.d.f. is obtained from Gamma integral, the distribution is reffered to as Gamma Distribution.

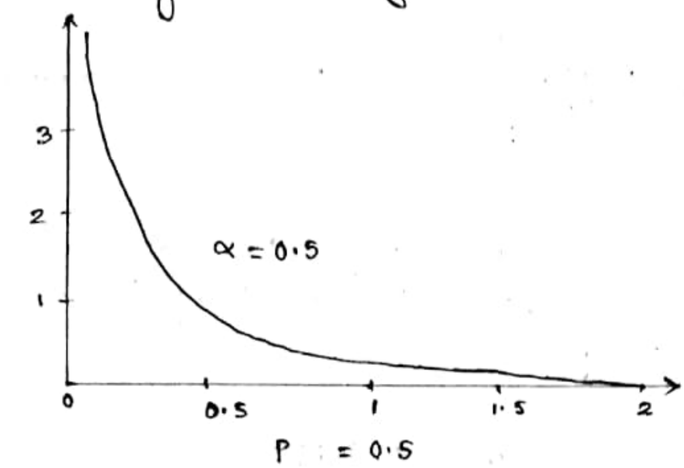
We denote the distn. by $G(\alpha, p)$ or $\gamma(\alpha, p)$.

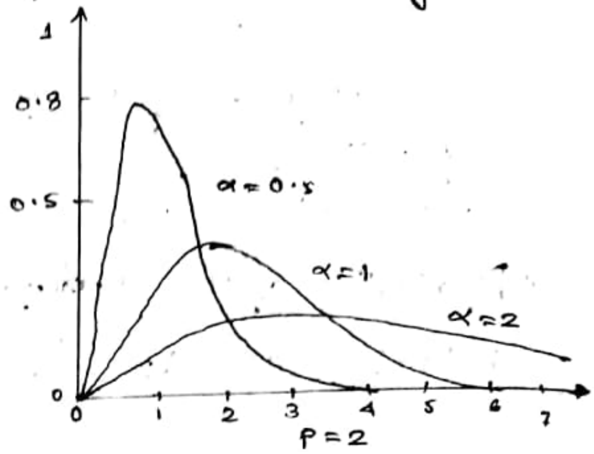
Clearly $\gamma(\alpha, 1)$ is an exponential distribution with mean α.

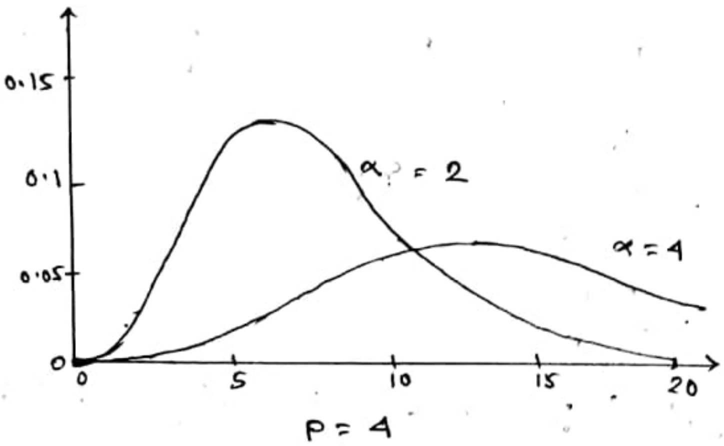
Note:– $\gamma(1, p)$ is known as a Standard Gamma Distribution, and we denote the distribution by $\gamma(p)$.

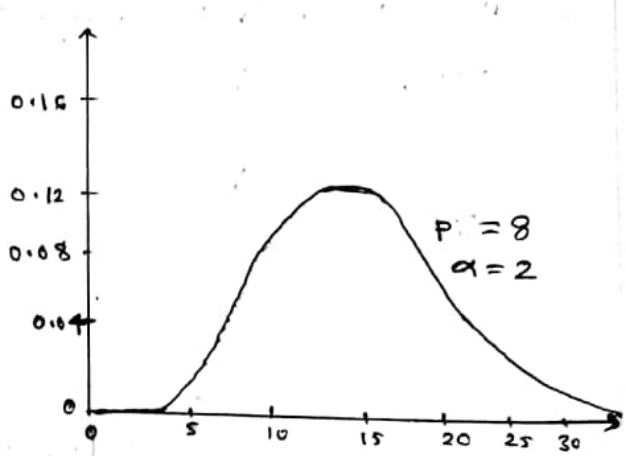
Properties :-

1. Gamma distribution is bell-shaped, positively skewed, leptokurtic distribution depending on the choice of the parameters ; it may be slightly skewed, in significantly leptokurtic or, it may be significantly skewed and leptokurtic. Here the term of the Gamma density can be thought of a slight or, significant departure from natural density.

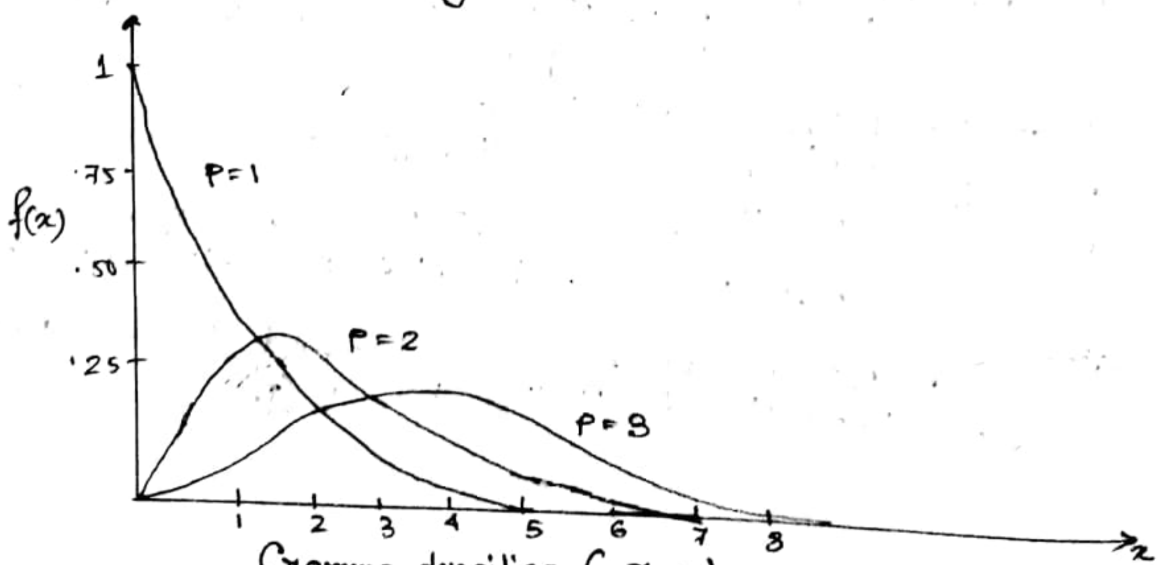








: Gamma Density function :



Gamma densities ($\alpha = 1$)

2. Additive Property of Gamma distribution :—

Suppose $X_1, \ldots, X_K$ are independently distributed gamma variables where

$$X_i \sim \gamma(\alpha, p_i) \quad ; \; i = 1(1)K.$$

Define, $S_K = \sum_{i=1}^{K} X_i$ then $S_K \sim \gamma(\alpha, \sum_{i=1}^{K} p_i)$.

■ Reproductive property :→ Proof:–

$$M_{S_K}(t) = \prod_{i=1}^{K} M_{X_i}(t) = \prod_{i=1}^{K} \left(1 - \frac{t}{\alpha}\right)^{-p_i}$$

$$= \left(1 - \frac{t}{\alpha}\right)^{-\sum_{i=1}^{K} p_i}$$

$$\therefore S_K \sim \gamma\left(\alpha, \sum_{i=1}^{K} p_i\right)$$

Note that the distribution function of gamma distribution can be expressed explicitly in terms of that of a suitable Poisson distribution.

■ Moments:—

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

$$= \int_0^\infty \frac{x^{r+p-1} \cdot e^{-\alpha x}}{\Gamma p} \cdot \alpha^p \, dx$$

$$= \frac{1}{\alpha^r} \cdot \frac{\Gamma(r+p)}{\Gamma(p)}$$

Putting, $r = 1$,

$$\mu_1' = E(X) = \frac{\Gamma(p+1)}{\alpha \Gamma p} = \frac{p}{\alpha}.$$

$r = 2$, $\mu_2' = E(X^2) = \frac{\Gamma(p+2)}{\alpha^2 \Gamma p} = \frac{p(p+1)}{\alpha^2}$.

$$\therefore Var(X) = \frac{p(p+1)}{\alpha^2} - \frac{p^2}{\alpha^2} = \frac{p}{\alpha^2}$$

* Thus in case of standard Gamma distribution where $\alpha = 1$, then Mean = Variance. *

■ Mode of the Distribution:—

$$f(x) = \frac{\alpha^p}{\Gamma(p)} e^{-\alpha x} \cdot x^{p-1}$$

$$\therefore \frac{d}{dx}[f(x)] = \frac{\alpha^p}{\Gamma(p)} \left[x^{p-1} \cdot (-\alpha) \cdot e^{-\alpha x} + e^{-\alpha x} \cdot (p-1) x^{p-2} \right]$$

$$= \frac{\alpha^p}{\Gamma(p)} \cdot x^{p-2} \cdot e^{-\alpha x} \left[-\alpha x + p - 1 \right] = 0$$

$$\therefore x = 0, \frac{p-1}{\alpha}$$

$\therefore x = \frac{p-1}{\alpha}$ where $p > 1$

$\therefore f'(x) > 0$ when $x < \frac{p-1}{\alpha}$

$\therefore f'(x) < 0$ when $x > \frac{p-1}{\alpha}$.

Scanned by CamScanner

Moment Generating Function :

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

$$= \int_0^\infty e^{tx} \cdot \frac{\alpha^p}{\Gamma p} \cdot e^{-\alpha x} \cdot x^{p-1}\,dx$$

$$= \frac{\alpha^p}{\Gamma p} \int_0^\infty e^{(t-\alpha)x} \cdot x^{p-1}\,dx$$

$$= \frac{\alpha^p}{\Gamma p} \int_0^\infty e^{-(\alpha-t)x} \cdot x^{p-1}\,dx$$

$$= \frac{\alpha^p \cdot \Gamma(p)}{(\alpha-t)^p \cdot \Gamma(p)}$$

$$= \left(1-\frac{t}{\alpha}\right)^{-p} \quad \text{when } |t|<\alpha.$$

$$= \sum_{k=0}^{\infty} \frac{(-p)_k}{k!}\left(-\frac{t}{\alpha}\right)^k$$

$$\therefore \mu_r' = \text{Coefficient of } \frac{t^r}{r!} = (-p)_r \left(-\frac{1}{\alpha}\right)^r$$

$$= (p+r-1)_r \cdot \frac{1}{\alpha^r}$$

$$\therefore \mu_1' = \frac{p}{\alpha} = \text{Mean.}$$

$$\therefore \mu_2' = \frac{p(p+1)}{\alpha^2}$$

$$\therefore \mu_2 = \frac{p(p+1)}{\alpha^2} - \frac{p^2}{\alpha^2} = \frac{p}{\alpha^2} = Var(X)$$

Cumulant Generating Function :

$$K(t) = \ln M(t)$$

$$= -p\ln\left(1-\frac{t}{\alpha}\right), \quad |t|<\alpha$$

$$= p\sum_{r=1}^{\infty} \frac{\left(\frac{t}{\alpha}\right)^r}{r} = p\sum_{r=1}^{\infty} \frac{t^r}{r!} \cdot \frac{(r-1)!}{\alpha^r}$$

$$K_r = \frac{p(r-1)!}{\alpha^r} = \text{coefficient of } \frac{t^r}{r!}$$

$$\text{Mean} = K_1 = \frac{p}{\alpha}\ ; \quad \text{variance} = K_2 = \frac{p}{\alpha^2} = \mu_2$$

$$\mu_3 = K_3 = \frac{2p}{\alpha^3}\ ; \quad \mu_4 = 3K_2^2 + K_4 = \frac{6p}{\alpha^4} + \frac{3p^2}{\alpha^4} = \frac{3p(2+p)}{\alpha^4}.$$

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2p \cdot \cancel{\alpha^3}}{\cancel{\alpha^3} \cdot p^{3/2}} = \frac{2}{\sqrt{p}} > 0$$

Here the distribution is positively skewed.

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6}{p} > 0$$

$\therefore$ The distribution is leptokurtic.

Scanned by CamScanner

Distribution function of a Gamma Random Variable :—

Let $X \sim G(\alpha, p)$.
The distribution function of x is given by, —

$$F(K) = P[X \le K]$$
$$= \int_0^K f(x)dx = \int_0^K \frac{\alpha^p}{\Gamma p} \cdot e^{-\alpha x} \cdot x^{p-1} dx$$

Now,
$$I_p = \int_0^K \frac{\alpha^p}{\Gamma p} \cdot e^{-\alpha x} \cdot x^{p-1} dx = \frac{\alpha^p}{\Gamma p}\Big[-\frac{1}{\alpha} \cdot e^{-\alpha x} \cdot x^{p-1} \Big|_0^K + \frac{p-1}{\alpha}\int_0^K e^{-\alpha x} \cdot x^{p-2} dx \Big]$$
$$= \frac{-(\alpha K)^{p-1}}{\Gamma p} \cdot e^{-\alpha K} + \frac{\alpha^{p-1}}{\Gamma p-1}\int_0^K e^{-\alpha x} \cdot x^{p-2} dx$$
$$= \frac{-(\alpha K)^{p-1} e^{-\alpha k}}{\Gamma p} + I_{p-1}$$
$$= -\frac{(\alpha k)^{p-1} e^{-\alpha k}}{\Gamma p} - \frac{(\alpha K)^{p-2} e^{-\alpha k}}{\Gamma p-1} \cdots\cdots - \frac{\alpha k \cdot e^{-\alpha k}}{\Gamma 2} + I_1$$
$$= -\sum_{i=1}^{p-1} \frac{(\alpha k)^i \cdot e^{-\alpha k}}{\Gamma i+1} + \int_0^K \alpha e^{-\alpha x} dx$$
$$= -\sum_{i=0}^{p-1} \frac{(\alpha k)^i e^{-\alpha k}}{i!} + 1$$

$$\therefore 1 - P[X \le K] = \sum_{i=0}^{p-1} \frac{(\alpha k)^i \cdot e^{-\alpha k}}{i!}$$

$$\therefore P[X > K] = \sum_{i=0}^{p-1} \frac{(\alpha k)^i \cdot e^{-\alpha k}}{i!} = P[Y \le p-1]$$, where $Y \sim P(\alpha k)$

☆ **Problem 1:**— If a random variable has a pdf $f(x) = \frac{1}{\Gamma\alpha \cdot \beta^\alpha} \cdot e^{-\frac{x}{\beta}} \cdot x^{\alpha-1}$

where $x \geq 0$ and $\alpha, \beta > 0$.

Show that, —

$E\{g(X)(X - \alpha\beta)\} = \beta \cdot E\{X g'(X)\}$, provided both the expectation exist.

ANS:— $E\{g(X)(X-\alpha\beta)\}$

$$= \int_0^\infty g(x)(x-\alpha\beta) \cdot \frac{1}{\Gamma\alpha \cdot \beta^\alpha} \cdot e^{-x/\beta} \cdot x^{\alpha-1} dx$$
$$= \int_0^\infty g(x) \cdot x^\alpha \cdot e^{-x/\beta} \cdot \frac{1}{\Gamma\alpha \cdot \beta^\alpha} dx - \alpha\beta E[g(X)]$$
$$= \frac{1}{\Gamma\alpha \cdot \beta^\alpha}\Big[g(x) \cdot x^\alpha \cdot e^{-x/\beta}(-\beta)\Big|_0^\infty + \beta\int_0^\infty e^{-x/\beta}\{\alpha x^{\alpha-1} g(x) + x^\alpha g'(x)\}\Big] - \alpha\beta E[g(X)]$$

$$= \frac{1}{\Gamma\alpha \cdot \beta^{\alpha}} \left[0 + \alpha\beta \int_0^\infty g(x) \cdot e^{-x/\beta} \cdot x^{\alpha-1} dx + \beta \int_0^\infty x g'(x) \cdot e^{-x/\beta} \cdot x^{\alpha-1} dx \right] - \alpha\beta E[g(x)]$$

$$= \cancel{\alpha\beta E\{g(x)\}} + \beta E(X g'(X)) - \cancel{\alpha\beta E\{g(x)\}}$$

$$= \beta E\{X g'(X)\}.$$

☆ **Problem 2.** Evaluate Harmonic Mean (HM) of the Geometric distribution.

Soln. →

$$\mu_r' = \frac{1}{\alpha^r} \cdot \frac{\Gamma(r+p)}{\Gamma(p)}, \quad X \sim G(\alpha, p).$$

$$HM = \frac{1}{E\left(\frac{1}{X}\right)} = \frac{1}{E(X^{-1})} = \frac{1}{\mu_{-1}'}$$

$$= \frac{1}{(\alpha)^{-1}} \frac{\Gamma(p-1)}{\Gamma(p)}$$

$$= \frac{\alpha}{p-1}. \quad \text{(Ans)}$$

ANOTHER FORM OF THE PDF OF GAMMA DISTRIBUTION: —

$$f(x) = \begin{cases} \frac{1}{\theta^n \Gamma n} \cdot e^{-x/\theta} \cdot x^{n-1} & \text{if } 0 < x < \infty \\ 0 & \text{ow} \end{cases}$$

where $\theta > 0$, $n > 0$, $X \sim \text{Gamma}(\theta, n)$

Derivation of the p.d.f.:— Considering the improper integral, $\int_0^\infty e^{-x/\theta} x^{n-1} dx$, it converges iff $\theta > 0$, $n > 0$.

Now, $\int_0^\infty e^{-x/\theta} x^{n-1} dx = \int_0^\infty e^{-z} \cdot z^{n-1} \cdot \theta^n dz$ $\quad z = \frac{x}{\theta} \Rightarrow x = \theta z$, $dz = \frac{1}{\theta} dx$

$$= \theta^n \Gamma n$$

$\Rightarrow \frac{1}{\theta^n \Gamma n} \int_0^\infty e^{-x/\theta} x^{n-1} dx = 1 \Rightarrow$ It is verified that it's a p.d.f.

The above integral converges for $\theta > 0$, $n > 0$

$\therefore$ So, $X \sim \gamma(\theta, n)$ or Gamma (θ, n).

Moments:— $\mu_r' = E(X^r) = \int_0^\infty x^r \cdot \frac{1}{\theta^n \Gamma n} \cdot e^{-x/\theta} x^{n-1} dx$

$$= \int_0^\infty \frac{e^{-x/\theta} x^{n+r-1}}{\theta^n \Gamma n} dx$$

$$= \frac{\theta^{n+r} \Gamma(n+r)}{\theta^n \Gamma n} \quad \left[\text{where } \Gamma n = (n-1)!\right]$$

$\therefore$ Mean $= \frac{\theta \Gamma(n+1)}{\Gamma n} = \theta n$, Variance $= \mu_2' - \mu_1'^2 = \frac{\theta^2 \Gamma(n+2)}{\Gamma n} - (\theta n)^2$

$$= \theta^2 n.$$

BETA DISTRIBUTION :

- **Beta Distribution of 1st & 2nd kind :** → The idea of probability density can easily be developed from following improper integral :

i) $\int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \beta(a,b)$; $a, b > 0$

ii) $\int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}}\,dx = \beta(a,b)$; $a, b > 0$

⟨i⟩ & ⟨ii⟩, respectively, lead to the following probability density :

i) $f_1(x) = \frac{1}{\beta(a,b)} \cdot x^{a-1}(1-x)^{b-1}\, I_x(0,1)$; $a,b>0$ ...... (1)

ii) $f_2(x) = \frac{1}{\beta(a,b)} \cdot \frac{x^{a-1}}{(1+x)^{a+b}} \cdot I_x(0,\infty)$; $a,b>0$ ...... (2)

Clearly, $a>0$, $b>0$

$f_1(x) \geq 0 \;\forall x$ & $f_2(x) \geq 0 \;\forall x$ &

$\int_0^1 f_1(x)\,dx = 1$ & $\int_0^\infty f_2(x)\,dx = 1$.

The probability distribution having pdf $f_1(x)$ is called Beta Distribution of 1st kind, we denote this distn. by $\beta_1(a,b)$ and the probability distribution with pdf $f_2(x)$ is called Beta distribution of 2nd kind, we denote the distn. by $\beta_2(a,b)$. Depending on the parameters a, b ; Beta distn. of the 1st kind may be uniform symmetric bell shaped, symmetric U shaped, positively skewed bell-shaped, positively skewed J-shaped, negatively skewed bell-shaped and negatively skewed J-shaped, etc. Depending upon the parameters some of the moment of Beta distribution of 2nd kind do not exist. Hence MGF of Beta distribution does not exist, condition for the existance of rth order raw moments can easily be obtained as follows, —

$$\mu_r' = \int_0^\infty \frac{x^{a+r-1}\,dx}{\beta(a,b)\,(1+x)^{a+b}} = \frac{1}{\beta(a,b)} \int_0^\infty \frac{x^{a+r-1}\,dx}{(1+x)^{(a+r)+(b-r)}}$$

$$= \frac{1}{\beta(a,b)}\,\beta(a+r,\, b-r)$$

[Provided, $a > -r$ and $b > r$]

If we consider the integers moments then the rth order moment exists if b exceeds r.

Note :-

if $X \sim \beta_2(a,b) \Leftrightarrow \frac{X}{1+X} \sim \beta_1(a,b)$

or, conversely,

if $X \sim \beta_1(a,b) \Leftrightarrow \frac{X}{1-X} \sim \beta_2(a,b)$.

$X \sim B_2(a,b)$

$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{B(a,b)} \int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}}\,dx$

$\frac{x}{1+x} = y \Rightarrow 1-y = \frac{1}{1+x}$
$\Rightarrow \frac{dy}{dx} = (1+x)^{-2}$

$\therefore \int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{B(a,b)} \int_0^1 y^{a-1}(1-y)^{b-1}\,dy$

$= \frac{B(a,b)}{B(a,b)} = 1$

Beta Distn. of 1st kind:

Definition: An absolutely continuous random variable X defined over $[0,1]$ is said to follow Beta distribution with parameters a and b if its p.d.f. is given by, —

$$f(x) = \begin{cases} \frac{1}{\beta(a,b)} \cdot x^{a-1}(1-x)^{b-1} & ; \; 0<x<1, \; a>0, b>0 \\ 0 & \text{ow} \end{cases}$$

Notation, $X \sim \beta_1(a,b)$ or simply $\beta(a,b)$.

Moments:

$$\mu_r' = E(X^r) = \int_0^1 x^r \cdot \frac{1}{\beta(a,b)} x^{a-1} \cdot (1-x)^{b-1} dx$$

$$= \frac{\beta(a+r,b)}{\beta(a,b)}, \quad r > -a$$

$$= \frac{\Gamma(a+r) \cancel{\Gamma b} \, \Gamma(a+b)}{\Gamma a \, \cancel{\Gamma b} \, \Gamma(a+b+r)} = \frac{(a+r-1)!\,(a+b-1)!}{(a+b+r-1)!\,(a-1)!}$$

Putting, $r=1$, $E(X) = \frac{a!\,(a+b-1)!}{(a+b)!\,(a-1)!} = \frac{a}{a+b}$.

Putting, $r=2$, $E(X^2) = \frac{(a+1)!\,(a+b-1)!}{(a+b+1)!\,(a-1)!} = \frac{a(a+1)}{(a+b)(a+b+1)}$

$$\therefore \text{Var}(X) = \frac{a(a+1)}{(a+b)(a+b+1)} - \frac{a^2}{(a+b)^2}$$

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

$$= \frac{ab}{(a+b)^2(a+b+1)}.$$

☆ **Problem 1.** If $X \sim \beta(a,b)$ then show that $V(X) \le \frac{1}{4}$.

Soln. → $V(X) = \frac{ab}{(a+b)^2(a+b+1)}$

$$\therefore \frac{1}{4} - V(X) = \frac{(a+b)^2(a+b+1) - 4ab}{4(a+b)^2(a+b+1)}$$

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

$$= \frac{(a+b)^3 + (a-b)^2}{4(a+b)^2(a+b+1)} \ge 0 \quad \text{as } a>0 \,\&\, b>0$$

$$\therefore V(X) \le \frac{1}{4}.$$

Scanned by CamScanner

☆ Problem 2. If $X \sim \beta(a,b)$; S.T. AM > HM.

Soln. →
$$E(X) = \frac{a}{a+b} \qquad E\left(\frac{1}{X}\right) = \frac{(a-2)!\,(a+b-1)!}{(a+b-2)!\,(a-1)!} = \frac{a+b-1}{a-1}.$$

$$\therefore \frac{1}{E\left(\frac{1}{X}\right)} = \frac{a-1}{a+b-1}, \quad a>1$$

$$\therefore E(X) - \frac{1}{E\left(\frac{1}{X}\right)} = \frac{a}{a+b} - \frac{a-1}{a+b-1}$$

$$= \frac{b}{(a+b)(a+b-1)} > 0 \quad [\because a>1]$$

$$\therefore E(X) > \frac{1}{E\left(\frac{1}{X}\right)}$$

$\therefore$ AM > HM.

☆ Problem 3. If $X \sim \beta(a,b)$. find the geometric mean.

Soln. → Let, geometric mean = g

$$\therefore \ln g = \int_0^1 \ln x . f(x)\,dx = \int_0^1 \ln x . \frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}\,dx$$

Now, $$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

$$\frac{\partial \beta(a,b)}{\partial a} = \int_0^1 x^{a-1}(1-x)^{b-1} \ln x\,dx$$

$$\therefore \ln g = \frac{1}{\beta(a,b)} \cdot \frac{\partial \beta(a,b)}{\partial a}$$

$$= \frac{\partial}{\partial a} \ln \beta(a,b)$$

$$= \frac{\partial}{\partial a} \ln \frac{\Gamma a\, \Gamma b}{\Gamma a+b}$$

$$= \frac{\partial}{\partial a}\left[\ln \Gamma a + \ln \Gamma b - \ln \Gamma a+b\right]$$

$$= \frac{\partial}{\partial a} \ln \Gamma a - \frac{\partial}{\partial a} \ln \Gamma a+b. \quad \text{(ANS)}$$

☆ **Problem 4.** If $X \sim \beta(a,b)$ then show that →

$$E\left[\left\{b - \frac{(a-1)(1-X)}{X}\right\} g(X)\right] = E\left[(1-X)\, g'(X)\right]$$

ANS:-

$$E\left[\left\{b - \frac{(a-1)(1-X)}{X}\right\} g(X)\right]$$

$$= \int_0^1 \left[b - \frac{(a-1)(1-x)}{x}\right] g(x)\, f(x)\, dx$$

$$= bE[g(X)] - \left[\int_0^1 (a-1)\, g(x) . x^{a-2} (1-x)^b . \frac{1}{\beta(a,b)} dx\right]$$

$$= bE[g(X)] - \left[(a-1) . \frac{x^{a-1}}{(a-1)} . g(x)(1-x)^b \Big|_0^1 - \int_0^1 x^{a-1}\left[(1-x)^b g'(x) - b(1-x)^{b-1} g(x)\right]\right] \frac{1}{\beta(a,}$$

$$= bE[\cancel{g(X)}] + \int_0^1 g'(x)(1-x) \frac{1}{\beta(a,b)} . x^{a-1}(1-x)^{b-1} dx - bE[\cancel{g(X)}]$$

$$= E[(1-X)\, g'(X)]$$

▧ **MGF of Beta Distribution :** Let, $X \sim \beta(a,b)$. The MGF of X is given by,

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

$$= E\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right)$$

since $0 < x < 1$, with probability 1.

Now,
$$\left|\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right| < \sum_{n=0}^{\infty} \frac{|t|^n}{n!} = e^{|t|}$$

$\therefore \sum_{n=0}^{\infty} \frac{(tX)^n}{n!}$ is a convergent series.

and hence,
$$M_X(t) = \sum_{n=0}^{\infty} E\left[\frac{(tX)^n}{n!}\right]$$

$$= \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$$

$$= \sum_{n=0}^{\infty} \frac{t^n}{n!} \times \frac{\beta(a+n,b)}{\beta(a,b)}$$

$$= \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(\prod_{k=0}^{n-1} \frac{a+k}{a+b+k}\right)$$, which can't be expressed in a compact form.

Mode of Beta Distribution : Suppose $X \sim \beta(a,b)$, when $a>1$ and $b>1$; the density is unimodal with mode $\frac{a-1}{a+b-2}$, when $a \le 1$ and $b \le 1$ and $a+b<2$ the density is either unimodal with mode at 1 or 0; or U-shaped with modes at both 0 and 1, when $a=1, b=1$, all points in $[0,1]$ are modes.

Now, $f(x) = \frac{1}{\beta(a,b)} \cdot x^{a-1}(1-x)^{b-1}$

$$f'(x) = \frac{1}{\beta(a,b)}\left[(a-1)x^{a-2}(1-x)^{b-1} - x^{a-1}(b-1)(1-x)^{b-2}\right]$$

$$= \frac{1}{\beta(a,b)} \cdot x^{a-2}(1-x)^{b-2}\left[(a-1)(1-x)-(b-1)x\right]$$

$$f'(x)=0 \Rightarrow (a-1)(1-x)-(b-1)x=0 \quad \left[\because \frac{1}{\beta(a,b)} \cdot x^{a-2}(1-x)^{b-2} \neq 0\right]$$

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

$$f''(x) = \frac{1}{\beta(a,b)}\Big[(a-2)x^{a-3}(1-x)^{b-2}\{(a-1)-x(a+b-2)\}$$
$$-x^{a-2}(b-2)(1-x)^{b-3}\{(a-1)-x(a+b-2)\}$$
$$-x^{a-2}(1-x)^{b-2}(a+b-2)\Big]$$

$$\therefore f''\left(\frac{a-1}{a+b-2}\right) = \frac{1}{\beta(a,b)}\left[0-0-\left(\frac{a-1}{a+b-2}\right)^{a-2}\left(\frac{b-1}{a+b-2}\right)^{b-2}(a+b-2)\right]$$

$$\therefore f''(x)\Big|_{x=\frac{a-1}{a+b-2}} < 0 \quad \text{when} \quad a,b>1$$

i.e. $f(x)$ has a maxima at $x = \frac{a-1}{(a+b-2)}$.

$\therefore$ for $a, b>1$; the distribution is unimodal with mode at
$$x = \frac{a-1}{a+b-2}.$$

Now, $f''(x)=0$ at $x=0$ and 1.

When $a \le 1, b \le 1, a+b<2$

We have either
$a=1, b<1$
$a<1, b=1$
$a<1, b<1$

When, $a=1, b<1$, $f(x) = \frac{(1-x)^{b-1}}{\beta(a,b)}$; $0<x<1$

$$\therefore f'(x) = \frac{1}{\beta(a,b)}(b-1)(-1)\cdot(1-x)^{b-2}, \quad 0<x<1$$
$$>0$$

$\therefore f(x)$ is an increasing function of x. $\therefore$ the mode is at $x=1$.

When $a<1$, $b=1$, $f(x)=\dfrac{x^{a-1}}{\beta(a,b)}$

$$\therefore f'(x)=\frac{(a-1)x^{a-2}}{\beta(a,b)}<0$$

$\therefore f(x)$ is a decreasing function of x.

$\therefore$ the mode is at $x=0$.

When $a<1$, $b<1$,

$f(x)\to\infty$ when $x\to 0$ and $x\to 1$.

also $f(x)$ is minimum at $x=\dfrac{a-1}{a+b-2}$.

Hence, the distribution is U shaped with modes at $0, 1$; when $a, b=1$, $f(x)=1$, all points in $[0,1]$ are modes.

Distribution Function of Beta Distribution :-

Let $X\sim\beta(a,b)$.

$$F_X(K)=\int_0^K \frac{x^{a-1}(1-x)^{b-1}\,dx}{\beta(a,b)}=I_{a,b}$$

i) Beta Distn of 2nd kind :

- **Definition** : A r.v. X is said to have Beta distribution of second kind if its PDF is

$$f(x) = \frac{1}{\beta(a,b)} \cdot \frac{x^{a-1}}{(1+x)^{a+b}} \quad , \quad 0 < x < \infty \ , a > 0, \quad b > 0.$$

$$= 0 \qquad\qquad , \text{ ow}$$

We write, that $X \sim \beta_2(a,b)$

Remark: – If $X \sim \beta_2(a,b)$ then $Y = \frac{X}{1+X} \sim \beta_1(a,b)$.

Proof:-
$$F_Y(y) = P(Y \le y)$$
$$= P\left[\frac{X}{1+X} \le y\right]$$
$$= P\left[X \le \frac{y}{1-y}\right] \quad \text{if } 0<y<1$$
$$= F_X\left(\frac{y}{1-y}\right) \quad \text{if } 0<y<1$$

Differentiating w.r.t. y,

$$f_Y(y) = f_X\left(\frac{y}{1-y}\right) \cdot \frac{d}{dy}\left(\frac{y}{1-y}\right), \text{ if } 0<y<1$$

$$= \frac{1}{\beta(a,b)} \cdot \frac{\left(\frac{y}{1-y}\right)^{a-1}}{\left\{1+\frac{y}{1-y}\right\}^{a+b}} \cdot \frac{1}{(1-y)^2}$$

$$= \frac{1}{\beta(a,b)} \cdot y^{a-1}(1-y)^{b-1} \quad \text{if } 0<y<1$$

Hence, $Y \sim \beta_1(a,b)$.

Moments:–

$$\mu_r' = \frac{1}{\beta(a,b)} \int_0^\infty \frac{x^{r+a-1}}{(1+x)^{a+b}}\,dx$$

$$= \frac{\beta(a+r, b-r)}{\beta(a,b)}$$

$$= \frac{\Gamma(r+a)\Gamma(b-r)}{\Gamma(a)\Gamma(b)} \quad , \quad -a<r<b.$$

It can be shown that,

$$E(X) = \frac{a}{b-1}, \text{ if } b>1$$

$$Var(X) = \frac{a(a+b-1)}{(b-1)^2(b-2)} \quad \text{if } b>2$$

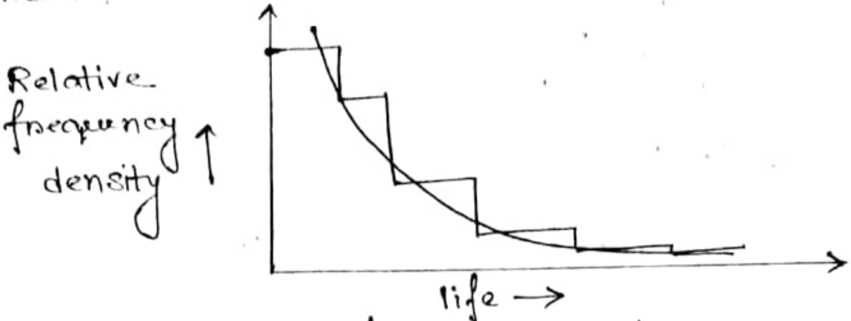
$$HM = \frac{a-1}{b} \quad \text{if } a>1.$$

Note that, moment μ_r', $r \ge b$, does not exist and consequently the MGF does not exist.

EXPONENTIAL DISTRIBUTION :

• One parameter Exponential Distribution :

Derivation of the pdf :→ Suppose we are given the distribution of life in hours of thousand dry cells of a particular brand. Here the observed distribution will be, —



Note that, here the frequency density, —

$$f_x \propto e^{-\theta x}, \text{ where } \theta > 0 \text{ determines the law explicitly.}$$

Therefore, it is reasonable to assume a probability density function, $f(x) \propto e^{-\theta x} I_x(0,\infty)$

If we regard average life to be μ then,

$$k\int_0^\infty x e^{-\theta x} dx = \mu \quad ; \quad \text{note that } \int_0^\infty k.e^{-\theta x} dx = 1 \text{ [as it's a p.d.f.]}$$

$$\Rightarrow k = \theta$$

$$\Rightarrow \theta \int_0^\infty x.e^{-\theta x} dx = \mu$$

$$\Rightarrow \frac{\Gamma 2}{\theta} = \mu$$

$$\Rightarrow \theta = \frac{1}{\mu}.$$

$$\therefore \boxed{f(x) = \frac{1}{\mu}.e^{-\frac{x}{\mu}} I_x(0,\infty)} \text{ ---- } (*)$$

It's the p.d.f. of one parameter Exponential distribution. We write, $X \sim E(\mu)$; where μ is the mean of the distribution.

Note:- Here we see that the distribution is 'J' shaped, positively skewed, thus we don't find any measure of kurtosis. We also see that, mean $(=\mu)$ of the distribution characterises the location as well as the spread of the distribution since in case of Exponential distribution.

• **Definition :—** An absolutely continuous random variable is said to follow exponential distn if the p.d.f. is of the form, —

$$f(x) = \frac{1}{\mu}.e^{-x/\mu} \quad ; \text{ where } 0 < x < \infty, \ \mu > 0.$$

We write — $X \sim E(0,\mu)$ or $E(\mu)$.

Moment generating function : →

$$M(t) = MGF = \frac{1}{\mu}\int_0^\infty e^{tx - \frac{x}{\mu}}\,dx = E(e^{tX})$$

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

[exists, if $|t| < \frac{1}{\mu}$]

$$= \frac{1}{\cancel{\mu}} \times \frac{\cancel{\mu}}{1-\mu t}$$

$$= (1-\mu t)^{-1}$$

$$= \sum_{j=0}^{\infty} \mu^j t^j \quad [\because |t| < \tfrac{1}{\mu}]$$

$\therefore \mu_j' = j!\,\mu^j$

$\therefore \mu_j' =$ coefficient of $\frac{t^j}{j!}$ in the expansion of $M(t)$.

$\therefore \mu_1' = \text{mean} = \mu$

$\therefore \mu_2' = 2\mu^2$; $\therefore V(X) = 2\mu^2 - \mu^2 = \mu^2$

Cumulant Generating function : →

$$K(t) = \ln M(t)$$

$$= \ln(1-\mu t)^{-1}$$

$$= -\ln(1-\mu t)$$

$$= \sum_{j=1}^{\infty} \frac{\mu^j t^j}{j}$$

$\therefore k_r =$ coefficient of $\frac{t^r}{r!}$

$= \mu^r (r-1)!$

$\therefore k_1 = \mu = \text{mean},$

$\therefore k_2 = \mu^2 = \text{variance},$ s.d. $= \mu$.

$\therefore k_3 = \mu_3 = 2\mu^3$

$\therefore k_4 = 6\mu^4$

$\therefore \mu_4 = k_4 + 3k_2^2 = 9\mu^4$.

CDF :- $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x \frac{1}{\mu} e^{-t/\mu}\,dt = 1 - e^{-x/\mu}$, if $x \geq 0$

$= 0$ ow

Scanned by CamScanner

* Exponential Distribution lacks memory:—

Loss of memory property:— Iff $X \sim$ exponential with mean μ then $P[X > x+y / X > x] = P[X > y]$, where $f(x) = \frac{1}{\mu} \cdot e^{-x/\mu}$; $0 < x < \infty$.

Proff:—

If part: → If $X \sim E(\mu)$ then, $P[X > x+y / X > x] = P[X > y]$
$\Rightarrow P[X > x+y] = P[X > x]\,P[X > y]$

• **ANS:**—

$$P[X > y] = \int_y^\infty \frac{1}{\mu} \cdot e^{-x/\mu}\,dx$$

$$= \frac{1}{\mu}\left[\frac{e^{-x/\mu}}{-1/\mu}\right]_y^\infty$$

$$= \frac{1}{\mu}\left[0 + \mu e^{-y/\mu}\right]$$

$$= e^{-y/\mu}$$

$$\therefore P[X > x+y] = e^{-(x+y)/\mu};$$

$$\therefore P[X > x] = e^{-x/\mu}$$

$$\therefore \frac{P[X > x+y]}{P[X > x]} = \frac{e^{-(x+y)/\mu}}{e^{-x/\mu}} = e^{-y/\mu} = P[X > y]$$

$$\Rightarrow P[X > x+y / X > x] = P[X > y].$$

* **NOTE:** Exponential distribution is the only continuous distribution which lacks memory.*

Only if Part: → If $P[X > x+y] = P[X > x]\,P[X > y]$, then $X \sim E(\mu)$.

• **ANS:**- $P[X > x+y+z] = P[X > x]\,P[X > y+z] = P[X > x]\,P[X > y]\,P[X > z]$

Let, $P[X > x] = \phi(x)$

it could be written that,— $\phi(\sum x) = \prod \phi(x)$

$$\therefore \phi(1) = \phi\left[\sum_{i=1}^{t} \frac{1}{t}\right] = \prod_{i=1}^{t} \phi\left(\frac{1}{t}\right) = \phi^t\left(\frac{1}{t}\right)$$

$$\therefore \phi\left(\frac{1}{t}\right) = \phi^{1/t}(1)$$

$$\therefore \phi(x) = \phi^{x}(1)$$

$$\Rightarrow \ln \phi(x) = x \ln \phi(1)$$

$$\Rightarrow \phi(x) = e^{-x(-\ln \phi(1))} = e^{-\lambda x} \quad [\text{where } \lambda = -\ln \phi(1)]$$

$$\therefore \int_x^\infty f(x)\,dx = e^{-\lambda x}$$

Differentiating both side w.r.t. x we get,—

$$\therefore 0 - f(x) = -\lambda e^{-\lambda x} \quad \therefore f(x) = \lambda e^{-\lambda x}$$

$$\therefore X \sim E\left(\frac{1}{\lambda}\right); \text{ where mean} = \frac{1}{\lambda}.$$

Note: → 1. $f(x) = \frac{1}{\mu} \cdot e^{-x/\mu}$

$$\therefore \phi(1) = \int_1^\infty \frac{1}{\mu} \cdot e^{-x/\mu} dx = \frac{1}{\mu}\left[-\mu \cdot e^{-x/\mu}\right]_1^\infty$$
$$= \left[e^{-1/\mu}\right]$$
$$\therefore \ln \phi(1) = -\frac{1}{\mu}.$$
$$\therefore \mu = -\frac{1}{\ln \phi(1)} = \frac{1}{\lambda} = \text{mean}.$$

2. If X be a non-negative continuous random variable satisfying the loss of memory property then X must be a exponential random variable. This is a characterisation of exponential distribution.

Mean Deviation about mean: —

$$MD_\mu(X) = \frac{1}{\mu} \int_0^\infty |x-\mu| \cdot e^{-x/\mu} dx$$
$$= \frac{2}{\mu} \int_\mu^\infty (x-\mu) \cdot e^{-x/\mu} dx$$

[let, $x-\mu = t$ & $dx = dt$]

$$= \frac{2}{\mu} \int_0^\infty t \cdot e^{-(t+\mu)/\mu} dt$$
$$= \frac{2}{\mu} \int_0^\infty t \cdot e^{-t/\mu} \cdot e^{-1} dt$$
$$= \frac{2}{e\mu} \int_0^\infty t \cdot e^{-t/\mu} dt$$
$$= \frac{2\mu}{e}$$
$$\therefore \frac{MD_\mu(X)}{s.d.} = \frac{2\mu}{e} \times \frac{1}{\mu}$$
$$= 2 e^{-1};$$

⇨ Definition :– A r.v. X is defined to have an (negative) exponential distribution with parameter $\mu\ (>0)$ if its pdf is given by

$$f(x, \mu) = \begin{cases} \frac{1}{\mu} \cdot e^{-x/\mu} & ; 0 < x < \infty, \mu > 0 \\ 0 & \text{ow} \end{cases}$$

The DF is, —

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ 1 - e^{-x/\mu} & , 0 \le x \end{cases}$$

Distribution of minimum of a set of independently distributed exponential variables :

Suppose $X_1, X_2, \ldots, X_n$ are independently distributed exponential variables.

Where, $X_i \sim$ exponential with mean θ_i

$$X_{(1)} = \text{Min}\{X_1, X_2, \ldots, X_n\}$$

$$\therefore P[X_{(1)} \le x] = 1 - P[X_{(1)} > x]$$

$$= 1 - P[X_1, X_2, \ldots, X_n > x]$$

$$= 1 - \prod_{i=1}^{n} P[X_i > x]$$

$$= 1 - \prod_{i=1}^{n} \left[\int_x^\infty \frac{1}{\theta_i} \cdot e^{-u/\theta_i} \, du \right]$$

$$= 1 - \prod_{i=1}^{n} e^{-x/\theta_i}$$

$$= \begin{cases} 1 - e^{-x \sum_{i=1}^{n} \frac{1}{\theta_i}} & ; x > 0 \\ 0 & \text{ow} \end{cases}$$

$$\therefore f_{X_{(1)}}(x) = \text{p.d.f. of } X_{(1)}$$

$$= \sum_{i=1}^{n} \frac{1}{\theta_i} \left[e^{-x \sum_{i=1}^{n} \frac{1}{\theta_i}} \right], \quad x > 0$$

Thus $X_{(1)}$ has exponential distribution with mean $\dfrac{1}{\sum_{i=1}^{n} \frac{1}{\theta_i}}$.

If X_i's are i.i.d. exponential variables with mean unity then $X_{(1)} \sim$ exponential with mean $\frac{1}{n}$.

☆ **Problem 1.** If $X \sim E(0, \theta)$ then show that, $P[X > r+s / X > r] = P[X > s]$, $r, s \in \mathbb{R}^+$.

ANS:- $P[X \le x] = 1 - e^{-x/\theta}$

$$\Rightarrow P[X > x] = e^{-x/\theta}$$

$$\therefore P[X > r] = e^{-r/\theta}$$

$$\therefore P[X > s] = e^{-s/\theta}$$

$$\therefore P[X > r+s] = e^{-(r+s)/\theta} = e^{-r/\theta} \cdot e^{-s/\theta} = P[X > r] \, P[X > s]$$

$$\therefore \frac{P[X > r+s]}{P[X > r]} = P[X > s]$$

$$\Rightarrow P[X > r+s \mid X > r] = P[X > s].$$

Scanned by CamScanner

☆ **Problem 2.** If $X \sim E(0, \theta)$, derive a recursion relation connecting central moments.

Soln → Mean $= \theta$,

$$\mu_r = E(X-\theta)^r$$

$$= \int_0^\infty (x-\theta)^r \cdot \frac{1}{\theta} \cdot e^{-\frac{x}{\theta}} dx$$

$$= \frac{1}{\theta} \int_0^\infty (x-\theta)^r \cdot e^{-x/\theta} dx$$

Differentiating w.r.t. θ, we get, —

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

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

$$= -r\mu_{r-1} + \frac{1}{\theta^2} \mu_{r+1}$$

$$\boxed{\therefore \mu_{r+1} = \theta^2 \left[\frac{d\mu_r}{d\theta} + r\mu_{r-1} \right]}$$

Putting, $r=1$, $\mu_2 = \theta^2 [1+0] = \theta^2$

$r=2$, $\mu_3 = \theta^2 [2\theta + 2\times 0]$
$= 2\theta^3$

$r=3$, $\mu_4 = \theta^2 [6\theta^2 + 3.\theta^2]$
$= 9\theta^4$.

— × —

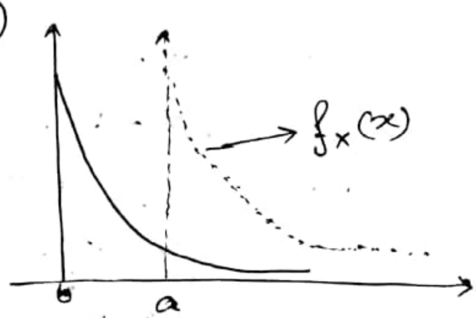
• Truncated Exponential / Shifted Exponential Distribution :

Derivation of the pdf : → Suppose, X is exponentially distributed with mean θ and we are to obtain the distribution of X given that $X > \alpha$, the p.d.f. of the truncated distribution is, —

$$g(x) = \frac{\frac{1}{\theta} \cdot e^{-x/\theta}}{\int_{\alpha}^{\infty} \frac{1}{\theta} \cdot e^{-x/\theta} dx} I_x(\alpha, \infty)$$

$$= \frac{\frac{1}{\theta} e^{-x/\theta}}{e^{-\alpha/\theta}} I_x(\alpha, \infty)$$

$$= \frac{1}{\theta} \cdot e^{-(x-\alpha)/\theta} I_x(\alpha, \infty).$$



Chossing $\theta = b\,(>0)$, $\alpha = a$, then, p.d.f. is ——

$$f(x) = \frac{1}{b} e^{-\frac{(x-a)}{b}} \quad ; \quad a < x < \infty, \; b > 0$$

• **Definition :—** An absolutely continuous random variable X defined over $[a, \infty)$ where $a \in \mathbb{R}$ is said to follow exponential distribution with parameter a and b if its pdf is given by, —

$$f(x) = \begin{cases} \frac{1}{b} \cdot e^{-\left(\frac{x-a}{b}\right)} & ; \; a \le x < \infty, \; b > 0 \\ 0 & \text{ow} \end{cases}$$

We write $X \sim E(a, b)$.

Particular case :— When $a = 0$, $b = \mu = \text{mean}\,(>0)$, then

$$f(x) = \frac{1}{\mu} \cdot e^{-x/\mu} \quad ; \quad 0 \le x < \infty$$

which is the p.d.f. of one parameter exponential distribution.

$$X \sim E(0, \mu).$$

Moments :— The rth order raw moment of X about a is given by

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

$$= \int_a^{\infty} (x-a)^r \cdot \frac{1}{b} \cdot e^{-\left(\frac{x-a}{b}\right)} dx \quad [\text{where } a \le x < \infty, \; b > 0]$$

[Let, $x - a = z$, $dx = dz$]

$$= \frac{1}{b} \int_0^{\infty} z^r \cdot e^{-z/b} dz$$

$$= b^r \, \Gamma(r+1)$$

$$= b^r \cdot (r!)$$

* It is important to note that Shifted exponential distn. does not have the "<u>Lack of memory Property</u>".

Putting, $r=1$, $E(X-a)=b$

$\Rightarrow E(X)=a+b$

Putting, $r=2$, $E(X-a)^2=2b^2$

$\Rightarrow E(X^2)-2aE(X)+a^2=2b^2$

$\Rightarrow E(X^2)=2b^2+a^2+2ab$

$\therefore Var(X)=\cancel{2}b^2+\cancel{a^2}+\cancel{2ab}-\cancel{a^2}-\cancel{2ab}-\cancel{b^2}$

$\therefore V(X)=b^2.$

Putting, $r=3$, $E(X-a)^3=6b^3$

$\Rightarrow E(X^3)-3aE(X^2)+3a^2E(X)-a^3=6b^3$

$\Rightarrow E(X^3)=6b^3+3a^2b+6ab^2+a^3$

Moment Generating function: —

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

$$=\int_a^\infty e^{tx}\cdot\frac{1}{b}\cdot e^{-\left(\frac{x-a}{b}\right)}dx$$

$$=\frac{1}{b}\cdot e^{\frac{a}{b}}\int_a^\infty e^{(tbx-x)/b}\,dx$$

$$=\frac{1}{b}\cdot e^{\frac{a}{b}}\left[\frac{e^{-\frac{x(1-tb)}{b}}}{(tb-1)/b}\right]_a^\infty \qquad \left[\because 1-tb>0 \Rightarrow t<\frac{1}{b}\right]$$

$$=e^{\frac{a}{b}}\left[0-\frac{e^{-\frac{a(1-tb)}{b}}}{(tb-1)}\right]$$

$$=\frac{e^{at}}{1-tb}.$$

Central moment Generating function: —

$$M_{X-\mu}(t)=E\left(e^{t(X-\mu)}\right)$$

$$=e^{-\mu t}\cdot M_X(t)$$

$$=e^{-(a+b)t}\cdot\frac{e^{at}}{1-tb}$$

$$=\frac{e^{-bt}}{(1-tb)}$$

$$=\left[\sum_{j=0}^\infty\frac{(-bt)^j}{j!}\right]\left[\sum_{j=0}^\infty (bt)^j\right]$$

Cumulant Generating Function :—

$$k(t) = \ln M_X(t)$$
$$= \ln \frac{e^{ta}}{1-bt}$$
$$= at - \ln(1-bt)$$
$$= at + \sum_{j=2}^{\infty} \frac{(bt)^j}{j} + bt$$
$$= (a+b)t + \sum_{j=2}^{\infty} \frac{(bt)^j}{j!} \times (j-1)!$$

$\therefore k_1 = E(X) = \mu = a+b =$ mean

$\therefore k_2 =$ variance $= \mu_2 = b^2$

$\therefore k_3 = \mu_3 = 2b^3$

$\therefore k_4 = 6b^4$

$$\therefore \mu_4 = k_4 + 3k_2^2$$
$$= 6b^4 + 3b^4$$
$$= 9b^4.$$

Measures of Skewness :—

$$\therefore \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{2b^3}{b^3} = 2$$

$\therefore$ The distribution is positively skewed.

Measures of kurtosis :—

$$\therefore \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{9b^4}{b^4} - 3 = 6$$

$\therefore$ The distribution is leptokurtic.

Mode of the distribution :—

$$f(x) = \frac{1}{b} \cdot e^{-\left(\frac{x-a}{b}\right)} \quad ; \quad a \le x < \infty, \; b > 0$$

$$\therefore \frac{d}{dx} f(x) = -\frac{1}{b^2} \cdot e^{-\left(\frac{x-a}{b}\right)} < 0$$

$\therefore f(x)$ is a decreasing function of x.

$\therefore f(x)$ is maximum when x is minimum.

$\therefore$ The mode of the distribution is at $x = a$.

Mean Deviation about mean :—

If $X \sim E(a,b)$ then $\mu = a+b$.

$$MD_{\mu}(X) = E|X-\mu|$$

$$= \int_a^{\infty} |x-\mu| f(x)\,dx$$

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

$$= 2\int_{a+b}^{\infty} (x-a-b)\frac{1}{b}\cdot e^{-\left(\frac{x-a}{b}\right)}dx$$

[let, $\frac{x-a}{b} = z$ $\therefore dx = b\,dz$]

$$= 2\int_{1}^{\infty} (bz-b)\cdot e^{-z}\,dz$$

$$= 2b\int_1^{\infty} z e^{-z} dz - 2b\int_1^{\infty} e^{-z} dz$$

$$= 2b\left[-ze^{-z}\Big|_1^{\infty} + \int_1^{\infty} e^{-z}dz\right] - 2b\left[-e^{-z}\right]_1^{\infty}$$

$$= 2b[e^{-1}+e^{-1}] - 2b[e^{-1}]$$

$$= 2be^{-1}.$$

Quantiles :—

Let ξ_p denotes the pth quantile of X.

$$F(\xi_p) = p \quad \forall\ p \in (0,1)$$

Now, $F(x) = \int_a^x f(t)\,dt = \frac{1}{b}\int_a^x e^{-\left(\frac{t-a}{b}\right)}dt$

[Put, $\frac{t-a}{b} = z$ $\Rightarrow dt = b\,dz$]

$$= \frac{1}{b}\int_0^{\frac{x-a}{b}} e^{-z}\, b\,dz$$

$$= \left[e^{-z}\right]_0^{\frac{x-a}{b}}$$

$$= 1 - e^{-\left(\frac{x-a}{b}\right)}$$

$\therefore F(\xi_p) = p$

or, $1 - e^{-\left(\frac{\xi_p - a}{b}\right)} = p$

or, $e^{-\left(\frac{\xi_p - a}{b}\right)} = 1-p$

or, $-\frac{\xi_p - a}{b} = \ln(1-p)$

or, $\xi_p = a - b\ln(1-p)$

1st quartile at $p=\frac{1}{4}$, $\xi_{1/4} = a - b\ln\frac{3}{4} = a + b\ln\frac{4}{3}$

2nd quartile at $p=\frac{1}{2}$, $\xi_{1/2} = a + b\ln 2 =$ Median.

3rd quartile at $p=\frac{3}{4}$, $\xi_{3/4} = a + b\ln 4$.

Quartile Deviation:—

$$Q.D. = \frac{\xi_{3/4} - \xi_{1/4}}{2} = \frac{b\ln 4 - b\ln\frac{4}{3}}{2} = \frac{b\ln 3}{2}.$$

Note:- Here, $\frac{Q.D.}{s.d.} = \frac{\cancel{b}\ln 3}{2\cancel{b}} = \frac{\ln 3}{2}$.

☆ **Problem 1.** If $X \sim E(a,b)$ then S.T. mean > median > mode.

ANS:- If $X \sim E(a,b)$ then we have already calculated,

mean $= a+b$, median $= a+b\ln 2$, mode $= a$.

as $b>0$, so, $a+b>a \Rightarrow$ mean > mode. —— (1)

We know, $\ln 2 < 1$

$\Rightarrow a+b\ln 2 < a+b \Rightarrow$ mean > median. —— (2)

and $a+b\ln 2 > a$ as $b>0$, $\ln 2 > 0$.

$\Rightarrow$ median > mode. —— (3)

∴ Combining (1), (2), (3), we get, $\longrightarrow$ Mean > median > mode.

☆ **Problem 2.** If $X \sim E(0,\frac{1}{b})$. find the mean deviation of X about its median.

ANS:- Median $= a+b\ln 2$ when $X \sim E(a,b)$

Here $X \sim E(0,\frac{1}{b})$, so median $= \frac{1}{b}\ln 2 = \xi_{1/2}$

$$MD_{\xi_{1/2}}(X) = E\left|X - \frac{\ln 2}{b}\right|$$

$$= 2\int_m^\infty (x-m)\, b.e^{-xb}\,dx \qquad \left[\text{where } \frac{\ln 2}{b} = m\right]$$

$$= 2b\int_0^\infty z.e^{-(z+m)b}\,dz \qquad [\text{let, } x-m=z,\ dx=dz]$$

$$= 2be^{-mb}\int_0^\infty z.e^{-bz}\,dz$$

$$= 2\cancel{b}.e^{-mb}.\frac{\Gamma 2}{b^{\cancel{2}}}$$

$$= \frac{2}{b}.e^{-\frac{\ln 2}{b}\times b}$$

$$= \frac{\cancel{2}}{b}\times\frac{1}{\cancel{2}} = \frac{1}{b}.\ (\text{ANS})$$

☆ **Problem 8.** Let X be an absolutely continuous random variable with the distribution function $F(x)$ and pdf $f(x)$, $x>0$. Let, $\alpha(x) = \frac{f(x)}{1-F(x)}$, $x>0$

$\alpha(x)$ = Failure rate function or Hazard function.

Show that, $\alpha(x)$ = constant iff X follows exponential distn.

$\boxed{\text{C.U.}}$

Soln. →

If part: → Let X is an exponential random variable.

i.e. $X \sim E(0, \theta)$

$$\therefore f(x) = \frac{1}{\theta} e^{-x/\theta}$$

$$\therefore 1 - F(x) = e^{-x/\theta}$$

$$\therefore \alpha(x) = \frac{\frac{1}{\theta} \cdot e^{-x/\theta}}{e^{-x/\theta}} = \frac{1}{\theta}$$

Only if part: → Suppose, $\alpha(x) = \alpha$ (constant)

$$\therefore \frac{f(x)}{1-F(x)} = \alpha$$

or, $\frac{\frac{d}{dx}F(x)}{1-F(x)} = \alpha$

or, $-\frac{d}{dx} \ln(1-F(x)) = \alpha$

or, $\ln[1-F(x)] = -\alpha x + c$

or, $1 - F(x) = ke^{-\alpha x}$ [where, $k = e^c$ = constant]

or, $F(x) = 1 - ke^{-\alpha x}$

or, $f(x) = k\alpha . e^{-\alpha x}$

Now, $F(0) = 1 - k$

or, $k = 1$,

$$\therefore f(x) = \alpha e^{-\alpha x}$$

i.e. $f(x) = \frac{1}{\theta} \cdot e^{-x/\theta}$ $\left[\because \theta = \frac{1}{\alpha}\right]$

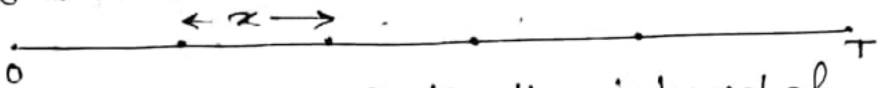
$\therefore X \sim E(0, \theta)$

Relationship between Poisson and Exponential distribution :→

- The length of time interval between occurance of two successive events can have an exponential distribution, provided that the no. of occurances in a fixed time interval follows Poisson distribution.

Proof:– Suppose that one of these occurences have just occured. To find the distribution of the length of time, say x, that one will have to wait until the next occurence.

Now, $P[X>t]$



= Probability that no occerences in the time interval of length t.

$= P[Y=0]$, where Y = the no. of occurences in the time interval of length 't' $\sim P(\mu t)$.

$$= e^{-\mu t} \cdot \frac{(\mu t)^0}{\lfloor 0} \quad , \text{if } t>0$$

$$= e^{-\mu t} \quad ; t>0$$

Hence the CDF of X, —

$$F_X(t) = P[X \leq t]$$
$$= 1 - P[X>t]$$
$$= 1 - e^{-\mu t}; \ t>0$$

The p.d.f. of X is, —

$$f_X(t) = \begin{cases} \mu e^{-\mu t} & \text{if } t>0 \\ 0 & \text{ow} \end{cases}$$

Therefore, X has an exponential distribution with mean $\frac{1}{\mu}$.

Remark:– 1) It can also be shown that if the inter-arrival times of occurances follow Exponential independently, then the number of occurances in a fixed time interval follows Poisson distribution.

2) Note the arithmetic :
Here μ is the mean rate of occurances of an event that is on an average μ occurances take place in 1 unit of time, i.e. one occurance take place in $\frac{1}{\mu}$ unit of time, i.e. the average or expected inter-arrival time is $\frac{1}{\mu}$.

Remark on HAZARD FUNCTION :—

Let T be the life-time of a component in a system. The probability of failure of the system in $[t, t+\Delta t]$ given that it has survived upto 't' is

$$P\left[t \leq T < t+\Delta t / T \geq t\right]$$

The instantaneous failure rate of the system at time point 't' is given by,

$$\lim_{\Delta t \to 0} \frac{P\left[t \leq T < t+\Delta t / T \geq t\right]}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\frac{P[t \leq T < t+\Delta t]}{\Delta t}}{P[T \geq t]}$$

$$= \lim_{\Delta t \to 0} \frac{\frac{F(t+\Delta t) - F(t)}{\Delta t}}{1-F(t)}$$

$$= \frac{F'(t)}{1-F(t)}$$

$$= \frac{f(t)}{1-F(t)}$$

Therefore, $\frac{f(t)}{1-F(t)}$ is called Hazard.

CDF of shifted exponential distribution.

ANS:- $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \begin{cases} \frac{1}{b} \int_a^x e^{-\left(\frac{t-a}{b}\right)} dt & \text{if } x > a \\ 0 & \text{ow} \end{cases}$

$= 1 - e^{-\left(\frac{x-a}{b}\right)}$ if $x > a$

$= 0$ ow.

DOUBLE EXPONENTIAL OR LAPLACE DISTRIBUTION : —

- **Derivation of the p.d.f. :→** Let us consider the mixture of the following densities :

$$f_1(x) = e^{x} I_x(-\infty, 0)$$

& $$f_2(x) = e^{-x} I_x(0, \infty)$$

Now, $f(x) = \alpha f_1(x) + (1-\alpha) f_2(x)$.

$$= \alpha e^{x} I_x(-\infty, 0) + (1-\alpha) e^{-x} I_x(0, \infty)$$

$\alpha \to$ mixing parameter, $\alpha \in (0, 1)$; let us choose $\alpha = \frac{1}{2}$, now, —

$$f(x) = \frac{1}{2}\left[e^{x} I_x(-\infty, 0) + e^{-x} I_x[0, \infty)\right]$$

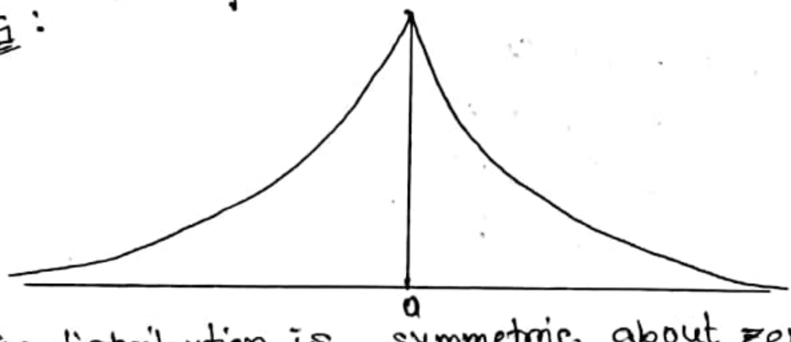
$$= \begin{cases} \frac{1}{2} e^{x} & \text{if } x < 0 \\ \frac{1}{2} e^{-x} & \text{if } x \geq 0 \end{cases}$$

$$= \frac{1}{2} e^{-|x|}, \quad -\infty < x < \infty.$$

The above probability density is in the standard form of the Laplace Distribution or, double exponential distribution.

Here the shape of the distribution is, —

PROPERTIES :



Clearly, the distribution is symmetric about zero.

i.e. the median = 0.

Further, it is unimodal, i.e. the mode = 0 [but $f(x)$ is not differentiable at zero.] . Again, the mean of the distribution exists. Hence

Mean = Median = Mode = 0

As the distn. is symmetric about zero, all odd ordered central moment vanishes.

$$F(x) = \int_{-\infty}^{x} \frac{1}{2} e^{-|u|} du$$

Moments:-

MGF:→ $E(e^{tX}) = \frac{1}{2}\int_{-\infty}^{\infty} e^{tx-|x|}\,dx$

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

$$= \frac{1}{2}\left[\int_{0}^{\infty} e^{-x(1+t)}\,dx + \int_{0}^{\infty} e^{-x(1-t)}\,dx\right]$$

$$= \frac{1}{2}\left[\frac{1}{1+t} + \frac{1}{1-t}\right]$$ [∵ 1st integral converges if $t > -1$, 2nd integral converges if $t < 1$; i.e. $|t| < 1$]

$$= (1-t^2)^{-1}$$

Central moment generating function about mean '0' = MGF

$$M_X(t) = \sum_{j=0}^{\infty} t^{2j} \quad \text{as } \mu = 0$$

$$\Rightarrow \mu_{2j} = (2j)!$$

$$\Rightarrow \mu_2 = 2,\ \mu_4 = 24$$

$$\therefore \mu_{2j} = 2j(2j-1)\mu_{2j-2},\ j = 1, 2, \ldots \quad (*)$$

[This is the Recursion relation for moments]

CGF:→ $K(t) = \ln M(t)$

$$= -\ln(1-t^2)$$

$$= \sum_{k=1}^{\infty} \frac{t^{2k}}{k}.$$

$K_1 = 0$
$K_2 = 2$
$K_3 = 0$
$K_4 = 12$

$\therefore \mu_2 = k_2 = 2$

$\therefore \mu_4 = K_4 + 3K_2^2 = 24.$

Measure of Kurtosis:-

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = 3$$

∴ The distribution is leptokurtic.

- **2 parameters Laplace Distribution :-**

- **Definition :-** An absolutely continuous random variable X defined over $(-\infty, \infty)$ is said to follow double exponential distribution with parameters μ and σ if its pdf is given by,—

$$f(x) = \frac{1}{2\sigma} \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|}, \quad -\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0 \quad \text{or,} \quad x \in \mathbb{R}, \; \mu \in \mathbb{R}, \; \sigma \in \mathbb{R}^+.$$

We denote or write $\rightarrow$ $X \sim DE(\mu, \sigma)$, i.e. $\frac{X-\mu}{\sigma} \sim DE(0,1)$

Particular case :- if $\mu = 0$ and $\sigma = 1$,

$$f(x) = \frac{1}{2} e^{-|x|}, \quad -\infty < x < \infty$$

Moments :-

$$E(X-\mu) = \frac{1}{2\sigma} \int_{-\infty}^{\infty} (x-\mu) \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|} dx$$

$$= \frac{1}{2\sigma} \int_{-\infty}^{\infty} \sigma z \cdot e^{-|z|} \cdot \sigma dz \qquad \left[\text{let, } \frac{(x-\mu)}{\sigma} = z\right]$$

$$= \frac{\sigma}{2} \int_{-\infty}^{\infty} z e^{-|z|} dz = 0 \qquad [\because \text{ the integrand is an odd function}]$$

$\therefore E(X) = \mu$.

The $(2r+1)$th order central moment of X is

$$\mu_{2r+1} = E(X-\mu)^{2r+1}$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\sigma} (x-\mu)^{2r+1} \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{2} (\sigma z)^{2r+1} \cdot e^{-|z|} dz$$

$$= 0 \qquad [\because \text{ the integrand is an odd function}]$$

$\therefore \mu_3 = 0$

$\therefore$ The coefficient of skewness is zero.

$\therefore$ The distribution is symmetric.

The all odd order central moments are zero here.

The $(2r)$th order central moment of X is, —

$$\mu_{2r} = E(X-\mu)^{2r}$$
$$= \int_{-\infty}^{\infty} \frac{1}{2\sigma}(x-\mu)^{2r} \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|} dx$$
$$= \frac{1}{2}\int_{-\infty}^{\infty} (\sigma z)^{2r} \cdot e^{-|z|} dz$$
$$= \frac{\cancel{2}}{\cancel{2}}\int_{0}^{\infty} \sigma^{2r} z^{(2r+1)-1} \cdot e^{-z} dz$$
$$= \sigma^{2r} \cdot \Gamma(2r+1)$$
$$= \sigma^{2r} \cdot (2r!)$$

Putting $r=1$, $\mu_2 = 2\sigma^2$ = variance.

$r=2$, $\mu_4 = 24\sigma^4$

$\therefore$ The coefficient of kurtosis, —

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{24\sigma^4}{4\sigma^4} - 3 = 3 > 0$$

$\therefore$ The distribution is leptokurtic.

NOTE: —

$$f(\mu+h) = \frac{1}{2\sigma} \cdot e^{-\left|\frac{\mu+h-\mu}{\sigma}\right|}$$
$$= \frac{1}{2\sigma} \cdot e^{-\left|\frac{h}{\sigma}\right|}$$
$$f(\mu-h) = \frac{1}{2\sigma} \cdot e^{-\left|\frac{h}{\sigma}\right|}$$
$$\therefore f(\mu+h) = f(\mu-h)$$

$\therefore$ The distribution is symmetric about μ.

Mean Deviation about mean: —

$$MD_\mu(X) = E|X-\mu| = \frac{1}{2\sigma}\int_{-\infty}^{\infty} |x-\mu| \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|} dx$$
$$= \frac{1}{2\cancel{\sigma}} \cdot \sigma \cancel{\sigma} \int_{-\infty}^{\infty} |z| \cdot e^{-|z|} dz \qquad \left[\text{let, } \frac{x-\mu}{\sigma} = z \quad \therefore dx = \sigma dz\right.$$
$$= \frac{\sigma}{\cancel{2}} \times \cancel{2} \int_0^{\infty} z e^{-z} dz$$
$$= \sigma \Gamma 2$$
$$= \sigma$$

$$\therefore \frac{MD}{SD} = \frac{1}{\sqrt{2}}.$$

QUANTILE :—

$$f(x) = \begin{cases} \frac{1}{2\sigma} \cdot e^{-\left(\frac{\mu - x}{\sigma}\right)} & \text{if } -\infty < x < \mu \\ \frac{1}{2\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)} & \text{if } \mu < x < \infty \end{cases}$$

The distribution function of X is,

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

if $x < \mu$,

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

if $x \geq \mu$,

$$F(x) = \frac{1}{2\sigma} \int_{-\infty}^{\mu} e^{-\frac{1}{\sigma}(\mu - t)} dt + \frac{1}{2\sigma} \int_{\mu}^{x} e^{-\frac{1}{\sigma}(t-\mu)} dt$$

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

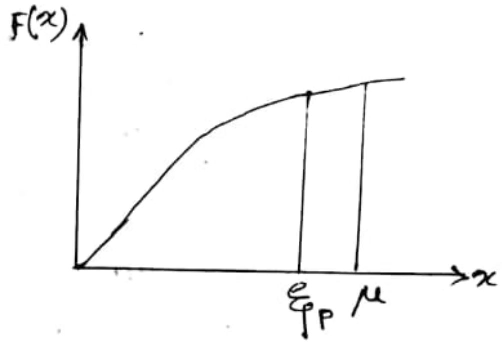
$$= \frac{1}{2} + \frac{1}{2}\left[1 - e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]$$

$$= 1 - \frac{1}{2} e^{-\frac{1}{\sigma}(x-\mu)}$$

$$\therefore F(x) = \begin{cases} \frac{1}{2} e^{-\frac{1}{\sigma}(\mu - x)} & , -\infty < x < \mu \\ 1 - \frac{1}{2} e^{-\frac{1}{\sigma}(x-\mu)} & , \mu < x < \infty . \end{cases}$$

Let ξ_p denotes the p^{th} quantile of X,

$$F(\xi_p) = p \quad \forall\ p \in (0,1)$$



if $p < \frac{1}{2}$ then $\xi_p < \mu$

if $p > \frac{1}{2}$ then $\xi_p > \mu$.

if $p < \frac{1}{2}$ then, $F(\xi_p) = p$

$$\Rightarrow \frac{1}{2} e^{-\frac{1}{\sigma}(\mu - \xi_p)} = p$$

$$\Rightarrow \frac{1}{\sigma}(\mu - \xi_p) = \ln 2p$$

$$\Rightarrow \xi_p = \mu + \sigma \ln 2p.$$

$$\therefore \xi_{1/4} = \mu + \sigma \ln 2 \cdot \frac{1}{4} = \mu - \sigma \ln 2$$

if $p > \frac{1}{2}$ the

$$F(\xi_p) = p$$

$$\Rightarrow 1 - \frac{1}{2} e^{-\frac{1}{\sigma}(\xi_p - \mu)} = p$$

$$\Rightarrow \frac{1}{\sigma}(\xi_p - \mu) = \ln 2(1-p)$$

$$\Rightarrow \xi_p = \mu - \sigma \ln 2(1-p).$$

$$\therefore \xi_{3/4} = \mu - \sigma \ln 2\left(1 - \frac{3}{4}\right) = \mu + \sigma \ln 2$$

$\therefore$ Quartile Deviation $= \frac{\xi_{3/4} - \xi_{1/4}}{2}$

$\therefore$ Q.D. $= \sigma \ln 2$.

Scanned by CamScanner

Mode of the Distribution:

$$f(x) = \begin{cases} \frac{1}{2\sigma} \cdot e^{-\frac{1}{\sigma}(\mu - x)} & , x < \mu \\ \frac{1}{2\sigma} \cdot e^{-\frac{1}{\sigma}(x-\mu)} & , x \geq \mu \end{cases}$$

Case-I for $x < \mu$,

$f'(x) = \frac{1}{\sigma} f(x) > 0$

Hence $f(x)$ is increasing function of x, when $x < \mu$.

$\therefore$ Here $f(x)$ is maximum when x is maximum.

Case-II for $x \geq \mu$,

$f'(x) = -\frac{1}{\sigma} f(x) < 0$

hence $f(x)$ is decreasing function of x, when $x \geq \mu$.

$\therefore$ Here $f(x)$ is maximum when x is minimum.

$\therefore$ Mode is at $x = \mu$.

Moment Generating Function (MGF):

$$M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{2\sigma} \cdot e^{-\left|\frac{x-\mu}{\sigma}\right|} dx$$

$$= \int_{-\infty}^{\mu} e^{tx} \cdot \frac{1}{2\sigma} \cdot e^{+\left(\frac{x-\mu}{\sigma}\right)} dx + \int_{\mu}^{\infty} e^{tx} \cdot \frac{1}{2\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)} dx$$

$$= I_1 + I_2.$$

$$I_1 = \frac{1}{2\sigma} \int_{-\infty}^{\mu} e^{tx} \cdot e^{-\frac{1}{\sigma}(\mu - x)} dx$$

$$= \frac{1}{2\sigma} \int_{-\infty}^{\mu} e^{tx - \frac{\mu}{\sigma} + \frac{x}{\sigma}} dx$$

$$= \frac{e^{-\frac{\mu}{\sigma}}}{2\sigma} \left[\frac{e^{x\left(t + \frac{1}{\sigma}\right)}}{t + \frac{1}{\sigma}} \right]_{-\infty}^{\mu} \qquad \left[t + \frac{1}{\sigma} > 0 \quad \therefore t > -\frac{1}{\sigma} \right]$$

$$= \frac{e^{-\frac{\mu}{\sigma}}}{2\sigma} \left[e^{\mu\left(t + \frac{1}{\sigma}\right)} \cdot \frac{1}{t + \frac{1}{\sigma}} \right]$$

$$= \frac{e^{\mu t}}{2\sigma\left(t + \frac{1}{\sigma}\right)} \quad , \text{when } t > -\frac{1}{\sigma}$$

Scanned by CamScanner

$$I_2 = \int_{\mu}^{\infty} e^{tx}\cdot\frac{1}{2\sigma}\cdot e^{-\left(\frac{x-\mu}{\sigma}\right)}dx$$

$$= \frac{1}{2\sigma}\int_{\mu}^{\infty} e^{\mu/\sigma}\cdot e^{x\left(t-\frac{1}{\sigma}\right)}dx$$

$$= \frac{e^{\mu/\sigma}}{2\sigma}\left[\frac{e^{x\left(t-\frac{1}{\sigma}\right)}}{\left(t-\frac{1}{\sigma}\right)}\right]_{\mu}^{\infty}$$

$$= -\frac{e^{\mu/\sigma}}{2\sigma}\cdot e^{\mu\left(t-\frac{1}{\sigma}\right)}\cdot\frac{1}{\left(t-\frac{1}{\sigma}\right)} \qquad \left[t-\frac{1}{\sigma}<0,\ \text{i.e. } t<\frac{1}{\sigma}\right]$$

$$= \frac{e^{\mu t}}{2\sigma\left(\frac{1}{\sigma}-t\right)}, \text{ when } t<\frac{1}{\sigma}.$$

$$\therefore \text{MGF} = \frac{e^{\mu t}}{2\sigma}\left[\frac{1}{t+\frac{1}{\sigma}}+\frac{1}{\frac{1}{\sigma}-t}\right] = \frac{e^{\mu t}}{2}\left[\frac{1}{1+\sigma t}+\frac{1}{1-\sigma t}\right]$$

$$= e^{\mu t}\left(1-t^2\sigma^2\right)^{-1}.$$

Central moment generating function:—

$$M_*(t) = E\left[e^{t(X-\mu)}\right] \quad \text{where } X\sim DE(\mu,\sigma).$$

$$= \int_{-\infty}^{\infty} e^{t(x-\mu)}\cdot\frac{1}{2\sigma}\cdot e^{-\left|\frac{x-\mu}{\sigma}\right|}dx \qquad \left[\text{let, } \frac{x-\mu}{\sigma}=z,\ \therefore dx=\sigma dz\right]$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} e^{t\sigma z}\cdot e^{-|z|}dz$$

$$= \frac{1}{2}\left[\int_{-\infty}^{0} e^{z(1+t\sigma)}dz + \int_{0}^{\infty} e^{-z(1-t\sigma)}dz\right]$$

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

$$= \left(1-t^2\sigma^2\right)^{-1} = \sum_{x=0}^{\infty}(t\sigma)^{2x} \qquad [\because |\sigma t|<1]$$

$$\therefore \mu_{2r} = \sigma^{2r}\cdot(2r)!$$

$$\therefore \mu_2 = 2\sigma^2,\ \mu_4 = 24\sigma^4.$$

As the distribution is symmetric so all odd order central moment vanishes.

$$\therefore \gamma_1 = 0 \text{ and } \gamma_2 = \frac{\mu_4}{\mu_2^2}-3 = 3.$$

Hence the distribution is leptokurtic.

$$M_*(t) = \left(1-t^2\sigma^2\right)^{-1}$$

$$\therefore M(t) = \text{MGF} = e^{\mu t}\cdot\left(1-t^2\sigma^2\right)^{-1}$$

$$\therefore \frac{d}{dt}[M(t)] = \frac{\mu e^{\mu t}}{1-t^2\sigma^2} + \frac{e^{\mu t}(+2t\sigma^2)}{1-t^2\sigma^2}$$

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

$$E\left[\frac{X-\mu}{\sigma}\right]=0$$

$$\therefore E(X)=\mu.$$

CDF:— $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

$$= \frac{1}{2\sigma}\int_{-\infty}^{x} e^{-\frac{|t-\mu|}{\sigma}}\,dt$$

If $x \le \mu$, then $F_X(x) = \frac{1}{2\sigma}\int_{-\infty}^{x} e^{\frac{t-\mu}{\sigma}}\,dt = \frac{1}{2}e^{\frac{x-\mu}{\sigma}}$.

If $x > \mu$, then $F_X(x) = \int_{-\infty}^{\mu} f_X(t)\,dt + \int_{\mu}^{x} f_X(t)\,dt$

$$= F_X(\mu) + \int_{\mu}^{x} \frac{1}{2\sigma}\cdot e^{-\frac{t-\mu}{\sigma}}\,dt$$

$$= \frac{1}{2} + \left[-\frac{1}{2}\cdot e^{-\frac{t-\mu}{\sigma}}\right]_{\mu}^{x}$$

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

Laplace's First & second Law of Error:→

In most of the prediction problem, following Laplace, the errors $e_i \overset{i.i.d.}{\sim} N(0,\sigma^2)$, $i=1(1)n$, and the PDF of $(e_1,\ldots,e_n)$ is $\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\sum_{i=1}^{n} e_i^2/2\sigma^2}$.

[Therefore, the least square principle is valid when $e_i \overset{i.i.d.}{\sim} N(0,\sigma^2)$.]

Otherwise, following Laplace, the errors $e_i \overset{i.i.d}{\sim} DE(0,\lambda)$ $\forall\ i=1(1)n$, and the PDF of $(e_1,\ldots,e_n)$ is $\left(\frac{1}{2\lambda}\right)^n \cdot e^{-\sum_{i=1}^{n}|e_i|/\lambda}$.

[in this case, better predicting formula can be obtained by minimizing $\sum_{i=1}^{n}|e_i|$.]

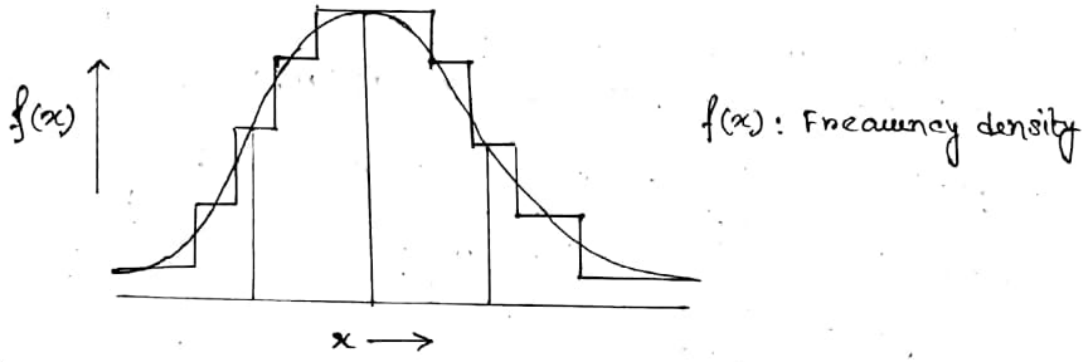
Therefore, Normal and Double exponential distribution are respectively known as Laplace's first law and second law of error.

Scanned by CamScanner

NORMAL DISTRIBUTION :

- Derivation of the p.d.f. :→

In several real life situation the frequency curve obtained by approximating the Histogram, takes the form like this :



Clearly,

1. The density curve is symmetric bell shaped (WLG let it be symmetric about '0').
2. The density function has points of inflations and asymtotes at both the tails.

Suppose X is a continuous random variable having pdf $f(x) \propto e^{-ax^2}$, $a>0$. In order to reflect faithfully the above features,

Suppose, $f(x) = c\,e^{-ax^2}$

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

$$\Rightarrow 2c \int_0^{\infty} e^{-ax^2} = 1$$

$$\Rightarrow c\int_0^{\infty} e^{-ay}.y^{\frac{1}{2}-1}dy = 1 \qquad \left[\text{let}, \to x^2 = y,\ 2x dx = dy \quad \therefore 2dx = \frac{dy}{x} = \frac{dy}{\sqrt{y}} \right]$$

$$\Rightarrow c = \sqrt{\frac{a}{\pi}} \qquad \left[\because \Gamma(\tfrac{1}{2}) = \sqrt{\pi} \right]$$

Clearly, $E(X) = 0$ [as if exists]

$$\sigma^2 = Var(X) = E(X^2) = c\int_{-\infty}^{\infty} x^2.e^{-ax^2}dx = 2c\int_0^{\infty} x^2.e^{-ax^2}dx$$

$$\therefore V(X) = c\int_0^{\infty} e^{-ay}y^{\frac{3}{2}-1}dy$$

$$\therefore \sigma^2 = c.\frac{\Gamma 3/2}{a^{3/2}} = \sqrt{\frac{a}{\pi}}.\frac{\frac{1}{2}\sqrt{\pi}}{a^{3/2}}$$

$$\therefore a = \frac{1}{2\sigma^2};\ c = \frac{1}{\sigma\sqrt{2\pi}} . \qquad \therefore V(X) = \frac{1}{2a} .$$

$$\therefore f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}}$$

Now, $X' = X + \mu$; X : random variable having pdf f.

Pdf of X' will be, —

$$P[X' = x] = g(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad x \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}^+.$$

This is the p.d.f. of the Normal distribution with mean μ and variance σ^2.

We denote the distribution by $X \sim N(\mu, \sigma^2)$.

- <u>Definition</u> :— An absolutely continuous random variable X defined over $(-\infty, \infty)$ is said to follow normal distribution with parameters μ, σ^2 if its pdf is given by, —

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

We write $X \sim N(\mu, \sigma^2)$.

☆ <u>Problem 1.</u> If $X \sim N(\mu, \sigma^2)$, then show that the distribution of X is symmetric about $x = \mu$.

<u>Soln.</u> :→

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad x \in \mathbb{R}, \ \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}^+.$$

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

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

$\therefore$ the distribution is symmetric about μ.

$$\therefore \xi_{1/2} = \mu.$$

<u>Moments</u> :—

$$E(x - \mu) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot (x - \mu)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} \cdot (\sigma z)\, dz$$

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

$\therefore \frac{1}{\sigma} dx = dz$

$$= 0 \quad [\because \text{It is an odd function}]$$

$\therefore E(X) = \mu$.

The $(2n+1)$th order central moment of X is

$$\mu_{2n+1} = E(X - \mu)^{2n+1} = 0$$

$\therefore$ All odd ordered central moments of normal distribution are zero.

$\therefore \mu_{2r} = E(X-\mu)^{2r}$

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

$$= \int_{-\infty}^{\infty} \sigma^{2r} z^{2r} \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} dz \qquad \left[\frac{x-\mu}{\sigma} = z \quad \therefore dx = \sigma dz \right]$$

$$= \frac{\sigma^{2r}}{\sqrt{2\pi}} \times 2 \int_{0}^{\infty} z^{2r} . e^{-z^2/2} dz$$

$$= \frac{\sigma^{2r}}{\sqrt{2\pi}} \times 2 \int_{0}^{\infty} (2u)^r . e^{-u} . \frac{du}{\sqrt{2u}} \qquad \left| \begin{array}{l} \frac{z^2}{2} = u \\ z\,dz = du \\ dz = \frac{1}{\sqrt{2u}} du \end{array} \right.$$

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

$$= \frac{\sigma^{2r} . 2^r}{\sqrt{\pi}} . \Gamma\left(r+\tfrac{1}{2}\right)$$

$$= \frac{\sigma^{2r} . 2^r}{\sqrt{\pi}} \left(r-\tfrac{1}{2}\right)\left(r-\tfrac{3}{2}\right) \cdots \tfrac{1}{2} . \Gamma\left(\tfrac{1}{2}\right)$$

$$= \sigma^{2r} (2r-1)(2r-3) \cdots 3 . 1 .$$

$$= \sigma^{2r} \binom{2r}{r} . \frac{r!}{2^r}$$

$$= \left(\frac{\sigma^2}{2}\right)^r . \binom{2r}{r} . r!$$

$$= \frac{(2r)!\, \sigma^{2r}}{r!\, 2^r} .$$

for $r=1$, $\mu_2 = \sigma^2$

$r=2$, $\mu_4 = 3\sigma^4$.

$\therefore$ Skewness $= \gamma_1 = 0$

$\therefore$ kurtosis $= \gamma_2 = \frac{3\sigma^4}{\sigma^4} - 3 = 0$

$\therefore$ Normal distribution is perfectly skewed and mesokurtic.

$X \sim N(\mu, \sigma^2)$

$$\therefore E(X-\mu)^{2r-1} = \int_{-\infty}^{\infty} (x-\mu)^{2r-1} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \int_{-\infty}^{\infty} (z\sigma)^{2r-1} . \frac{1}{\sqrt{2\pi}} . e^{-z^2/2} dz \quad , \text{ where } z = \frac{x-\mu}{\sigma}$$

$$= \frac{\sigma^{2r-1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2r-1} e^{-z^2/2} dz$$

$= 0$, since the integral $z^{2r-1} . e^{-z^2/2}$ is odd.

$\therefore$ All odd ordered central moment vanishes.

Moment Generating Function (MGF) :

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

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

$$= \int_{-\infty}^{\infty} \frac{e^{t(\mu+\sigma z)-\frac{1}{2}z^2}}{\sqrt{2\pi}} dz \qquad \text{let, } \frac{x-\mu}{\sigma} = z$$

$$= \frac{e^{\mu t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\sigma z - \frac{1}{2}z^2} dz$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z-\sigma t)^2} dz}_{} \quad \forall t \in \mathbb{R}$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad \left[\text{it is the pdf of } N(\sigma t, 1)\right]$$

i.e. $z \sim N(\sigma t, 1)$.

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

Central moment generating function :

$$M_*(t) = e^{-\mu t} \cdot M(t)$$

$$= e^{\frac{\sigma^2 t^2}{2}}$$

$$= \sum_{x=0}^{\infty} \left(\frac{\sigma^2 t^2}{2}\right)^x \cdot \frac{1}{x!}$$

$$= \sum_{x=0}^{\infty} \frac{t^{2x}}{(2x)!} \cdot \frac{(2x)!\,\sigma^{2x}}{2^x \cdot x!}$$

$$\therefore \mu_{2r} = \frac{(2r)!\,\sigma^{2r}}{2^r \cdot r!}$$

Cumulant Generating Function :

$$CGF = K(t) = \ln M(t) = \ln\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]$$

$$\therefore K(t) = \mu t + \frac{1}{2}\sigma^2 t^2$$

$\therefore$ $K_1 = \mu$

$\therefore$ $K_2 = \sigma^2$

$\therefore$ $K_3 = K_4 = \cdots = 0$

Recursion Relation for Central moments: —

$$\mu_{2r} = E(X-\mu)^{2r}$$
$$= \int_{-\infty}^{\infty} (x-\mu)^{2r} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Differentiating w.r.t. σ we get, —

$$\frac{d}{d\sigma}\mu_{2r} = \int_{-\infty}^{\infty} (x-\mu)^{2r} \cdot \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(-\frac{1}{\sigma^2}\right) + \frac{1}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(-\frac{1}{2}\right)(x-\mu)^2 \cdot \left(-\frac{2}{\sigma^3}\right)\right] dx$$

$$= -\frac{1}{\sigma}\mu_{2r} + \frac{1}{\sigma^3}\mu_{2r+2}$$

$$\therefore \mu_{2r+2} = \sigma^3 \left[\frac{d\mu_{2r}}{d\sigma} + \frac{1}{\sigma}\mu_{2r}\right]$$

Alt. method: —

$$\mu_{2r} = \frac{(2r)!\, \sigma^{2r}}{2^r \cdot r!}$$
$$= 1.3.5.\ldots(2r-1) \cdot \sigma^{2r}$$
$$= \{1.3.5.\ldots(2r-2)\,\sigma^{2r-2}\}\,\sigma^2(2r-1)$$
$$= (2r-1)\sigma^2 \cdot \mu_{2r-2}.$$

Thus,
$$\mu_2 = \sigma^2$$
$$\mu_4 = 3\sigma^4$$
$$\mu_6 = 15\sigma^6$$
$$\mu_8 = 105\,\sigma^8, \text{ etc.} \ldots$$

Recursion Relation for Raw moments: —

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\therefore \frac{d\mu_r'}{d\sigma} = \int_{-\infty}^{\infty} x^r \cdot \frac{1}{\sqrt{2\pi}} \left[-\frac{1}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \frac{1}{\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \left(-\frac{1}{2}\right)(x-\mu)^2\left(-\frac{2}{\sigma^3}\right)\right] dx$$

$$= -\frac{1}{\sigma}\mu_r' + \int_{-\infty}^{\infty} x^r \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \frac{1}{\sigma^3}(x^2 - 2\mu x + \mu^2)\, dx$$

$$= -\frac{1}{\sigma}\mu_r' + \frac{1}{\sigma^3}\mu_{r+2}' - \frac{2\mu}{\sigma^3}\mu_{r+1}' + \frac{\mu^2}{\sigma^3}\mu_r'.$$

Mean Deviation about mean: —

$$MD_\mu(X) = E|X-\mu|$$
$$= \int_{-\infty}^{\infty} \frac{|x-\mu|}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

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

$\therefore dz = \frac{1}{\sigma} dx$

$$= \int_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}} \cdot |z| \cdot e^{-z^2/2} dz$$
$$= \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty z \cdot e^{-z^2/2} dz$$
$$= \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty e^{-u} du.$$

let, $\frac{z^2}{2} = u$

$\therefore z\,dz = du$

$$= \sigma\sqrt{\frac{2}{\pi}}$$

Hence,
$var(|X-\mu|)$
$= E(X-\mu)^2 - E^2|X-\mu|$
$= \sigma^2 - \sigma^2 \cdot \frac{2}{\pi}$
$= \sigma^2\left(1-\frac{2}{\pi}\right)$

$\therefore \sqrt{\mu_2} = \text{s.d.} = \sigma;\quad \frac{MD}{\text{s.d.}} = \frac{\sigma\sqrt{\frac{2}{\pi}}}{\sigma} = \sqrt{\frac{2}{\pi}} < 1.$

☆ **Problem 2.** If $X \sim N(\mu, \sigma^2)$ then show that —

$$E\{(X-\mu)g(X)\} = \sigma^2 E[g'(X)].$$

Hence find the recursion relation for central moments.

Soln. → $E[(X-\mu)g(X)] = \int_{-\infty}^{\infty} (x-\mu) g(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

$$= g(x) \times 0 - \left[\int_{-\infty}^{\infty} g'(x) \left[\int_{-\infty} (x-\mu) \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \right] dx \right]$$

$$= -\int_{-\infty}^{\infty} g'(x)(-\sigma^2) \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \sigma^2 E[g'(X)] \quad \text{[Proved]}$$

Now, Let, $g(x) = (x-\mu)^{2n-1}$

$$g'(x) = (2n-1)(x-\mu)^{2n-2}$$

$$\therefore E(X-\mu)^{2n} = \sigma^2 E[(2n-1)(x-\mu)^{2n-2}]$$

$$\therefore E(X-\mu)^{2n} = \sigma^2(2n-1) E(X-\mu)^{2n-2}.$$

Mode & Point of Inflation :

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

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

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

$$f''(x) = -\left[\left(\frac{x-\mu}{\sigma^2}\right) f'(x) + f(x) \cdot \frac{1}{\sigma^2}\right]$$

$$= -\left[-\left(\frac{x-\mu}{\sigma^2}\right)^2 f(x) + f(x) \cdot \frac{1}{\sigma^2}\right]$$

$$= \frac{f(x)}{\sigma^2}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right]$$

Now, $f'(x) = 0 \Rightarrow -f(x)\left(\frac{x-\mu}{\sigma^2}\right) = 0$

$\Rightarrow x = \mu.$

$$f''(\mu) = -\frac{f(\mu)}{\sigma^2} < 0$$

$\therefore$ $f(x)$ has a maxima at $x = \mu$.

$\therefore$ Mode of the distribution $= \mu$.

Now, $f''(x) = 0 \Rightarrow \frac{f(x)}{\sigma^2}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right] = 0$

$\Rightarrow \left(\frac{x-\mu}{\sigma}\right)^2 = 1$

$\Rightarrow x = \mu \pm \sigma.$

$\therefore$ $f(x)$ has a point of inflation at $x = \mu \pm \sigma$.

Now, $f''(x) \downarrow 0$ whenever $x \to \uparrow \infty$ or, $x \downarrow -\infty$.

$(\because \sigma^2 < \infty \Rightarrow x^2 f(x) \downarrow 0$ as $x \uparrow \infty)$

$\therefore$ $f(x)$ has asymptotes at $x = \pm\infty$.

Example :– Let $X \sim N(\mu, \sigma^2)$ then show that $E|X-a| \geq \sigma\sqrt{\frac{2}{\pi}}$, for any a.

Soln. → Mean deviation is minimum when it is measured about the median. Here median $= \mu$.

$\therefore E|X-a| \geq E|X - Me(X)| = E|X-\mu|$

$\geq \sigma\sqrt{\frac{2}{\pi}}$ when $X \sim N(\mu, \sigma^2)$.

Reproductive Property :—

Let, $X_1 \sim N(\mu_1, \sigma_1^2)$

$X_2 \sim N(\mu_2, \sigma_2^2)$

X_1 and X_2 are independent random variable.

Define, $S = l_1X_1 + l_2X_2 + l_0$, where l_i's are known constant.

$$M_S(t) = E(e^{tS})$$
$$= E\left(e^{t(l_0 + l_1X_1 + l_2X_2)}\right)$$
$$= e^{tl_0}\, E\left(e^{tl_1X_1}\right) E\left(e^{tl_2X_2}\right) \quad \text{[Due to independence]}$$
$$= e^{tl_0}\, M_{X_1}(tl_1)\, M_{X_2}(tl_2)$$
$$= e^{tl_0} \cdot e^{\mu_1 tl_1 + \sigma_1^2 t^2 l_1^2 \cdot \frac{1}{2}} \cdot e^{\mu_2 tl_2 + \frac{1}{2}\sigma_2^2 t^2 l_2^2}$$
$$= e^{t(l_0 + l_1\mu_1 + l_2\mu_2) + \frac{t^2}{2}(l_1\sigma_1^2 + l_2\sigma_2^2)}$$
$$= e^{t\mu_S + \frac{t^2}{2}\sigma_S^2}$$

As MGF uniquely determines normal distribution.

$$S \sim N(\mu_S, \sigma_S^2)$$

i.e. $S \sim N(l_0 + l_1\mu_1 + l_2\mu_2,\ l_1^2\sigma_1^2 + l_2^2\sigma_2^2)$

□ In general, if $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1(1)n$. where, X_i's are independently distributed then,—

$$S_n = \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$$

If further, X_i's are identically distributed with common mean μ & common variance σ^2 then,—

$$S_n \sim N(n\mu, n\sigma^2)$$

and $\overline{X}_n = \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

□ Some specific cases :—

i) $X \sim N(\mu, \sigma^2) \Leftrightarrow X + c \sim N(\mu + c, \sigma^2)$; c = constant,

ii) $X \sim N(\mu, \sigma^2) \Leftrightarrow cX \sim N(c\mu, c^2\sigma^2)$; $c \neq 0$

[if $c = 0$, then $P[cX = 0] = 1$, cX will be a degenerate random variable]

iii) $X \sim N(\mu, \sigma^2) \Leftrightarrow -X \sim N(-\mu, \sigma^2)$

$\therefore$ if $\mu = 0$, $X \overset{D}{=} -X$ [identical distribution]

iv) $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$X_2 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

if $\mu = 0$, $X_1 + X_2 \overset{D}{=} X_1 - X_2$

- **Standard Normal Variate:**

X is said to be a standard normal variate if $X \sim N(0,1)$.

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X-\mu}{\sigma} \sim N(0,1).$$

Let, $Z = \dfrac{X-\mu}{\sigma}$, $\therefore Z \sim N(0,1)$

The distribution function of Z is

$$\begin{aligned}\Phi(z) &= P[Z \le z] \\ &= P[X-\mu \le z\sigma] \\ &= P[X \le \mu+\sigma z] \\ &= F_X(\mu+\sigma z).\end{aligned}$$

, where F_X denotes the distribution function of X.

$\therefore$ p.d.f. of Z is

$$\begin{aligned}\varphi(z) &= \frac{d}{dz}\Phi(z) \\ &= \frac{d}{dz}F_X(\mu+\sigma z) \\ &= \sigma f(\mu+\sigma z) \\ &= \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}.\end{aligned}$$

$$\therefore \varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} I_z(-\infty, \infty).$$

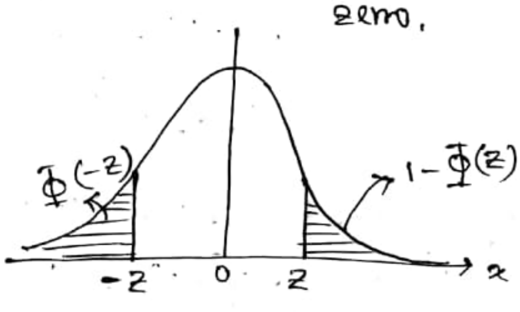
$$\therefore Z \sim N(0,1).$$

Properties of $\varphi(z)$ & $\Phi(z)$:

① $\varphi(z) = \varphi(-z)$, $\forall z$

② $\Phi(-z) = 1 - \Phi(z)$, $\forall z$

③ $\varphi(z)$ of $N(0,1)$ is symmetric about '0' and the mode is at zero.



Proof:-

$$\Phi(z) = \int_{-\infty}^{z} \varphi(u)\,du$$

$$\begin{aligned}\Phi(-z) &= \int_{-\infty}^{-z} \varphi(u)\,du \\ &= \int_{z}^{\infty} \varphi(u)\,du\end{aligned}$$

$$\begin{aligned}\therefore \Phi(z) + \Phi(-z) &= \int_{-\infty}^{z} \varphi(u)\,du + \int_{z}^{\infty} \varphi(u)\,du \\ &= \int_{-\infty}^{\infty} \varphi(u)\,du \\ &= 1 \quad \forall z\end{aligned}$$

Now, $\Phi(0) = \frac{1}{2}$.

Implication:- If $\Phi(z), z>0$ is given, then we can find $\Phi(-z)$ from the above formula.
$\therefore$ It is only necessary to provide the table of $\Phi(z)$ for positive values of z.

Note:-

$$P[X_1 > X_2]$$
$$= P[X_1 - X_2 > 0]$$
$$= P\left[\frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} > -\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right]$$
$$= P\left[Z > \frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right], \quad Z \sim N(0,1)$$
$$= 1 - \Phi\left(\frac{\mu_2 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$$

Now, if $\mu_1 = \mu_2$ then, $P[X_1 > X_2] = 1 - \Phi(0)$
$= 1/2$.

☆ **Problem 3.** If $X \sim N(0,1)$. find $E(\Phi(x))$ & $V(\Phi(x))$.

Ans:-
$$E(\Phi(x)) = \int_{-\infty}^{\infty} \Phi(x)\,\phi(x)\,dx = \int_0^1 z\,dz$$
[let, $\Phi(x) = z$
$\Rightarrow \phi(x)dx = dz$]

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

$$E(\Phi(x))^2 = \int_{-\infty}^{\infty} \{\Phi(x)\}^2 \phi(x)\,dx = \int_0^1 t^2 dt = \frac{1}{3}.$$

$$\therefore V(\Phi(x)) = \frac{1}{3} - \frac{1}{2^2} = \frac{1}{12}.$$

OR

$$Z = \Phi(x) \sim R(0,1)$$
$$\therefore E(\Phi(X)) = E(Z) = \int_0^1 z\,dz = \frac{1}{2}$$
$$E(Z^2) = \int_0^1 z^2 dz = \frac{1}{3}$$
$$\therefore V(Z) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

☆ **Problem 4.** If $X \sim N(0,1)$. find $E[X\Phi(X)]$ and $E[X^2\Phi(X)]$.

Ans:-
$$E(X\Phi(X)) = \int_{-\infty}^{\infty} x\,\Phi(x)\phi(x)\,dx = \int_{-\infty}^{\infty} \Phi(x)\, x\phi(x)\,dx$$
$$= \left[\Phi(x)(-\phi(x))\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \phi^2(x)\,dx$$
$$= 0 + \int_{-\infty}^{\infty} \frac{1}{2\pi} \cdot e^{-x^2} dx$$
$$= \frac{1}{2\sqrt{\pi}}$$

• Another way :— $E(X\Phi(X))$

$$= \int_{-\infty}^{\infty} x\,\Phi(x)\,\phi(x)\,dx$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{x} \phi(u)du\right)\phi(x)\,dx$$

$$= \int_{-\infty}^{\infty} \phi(u) \left(\int_{u}^{\infty} x\phi(x)dx\right) du \qquad -\infty < u < x < \infty$$

$$= \int_{-\infty}^{\infty} \phi(u) \left[\int_{u}^{\infty} x . e^{-x^2/2} . \frac{1}{\sqrt{2\pi}}\,dx\right] du$$

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

$$= \frac{2}{(\sqrt{2\pi})^2} \int_{0}^{\infty} e^{-u^2} du \qquad \left[\begin{array}{l} u^2 = z \\ 2u\,du = dz \end{array}\right]$$

$$= \frac{1}{2\pi} \int_{0}^{\infty} e^{-z} . z^{\frac{1}{2}-1} dz$$

$$= \frac{\Gamma 1/2}{2\pi} = \frac{\sqrt{\pi}}{2\pi} = \frac{1}{2\sqrt{\pi}} .$$

Now, $E(X^2\Phi(X)) = \int_{-\infty}^{\infty} x^2 \phi(x)\,\Phi(x)\,dx$

$$= \Big[-x\Phi(x)\,\phi(x)\Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \phi(x)\left[\Phi(x) + x\phi(x)\right] dx$$

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

$$= I_1 + I_2 .$$

$$I_1 = \int_{-\infty}^{\infty} \phi(x)\,\Phi(x)\,dx$$

$$= \int_0^1 z\,dz = \frac{1}{2} \qquad \left[\begin{array}{l} \text{let, } \Phi(x) = z \\ \phi(x)\,dx = dz \end{array}\right]$$

$$I_2 = \int_{-\infty}^{\infty} x\,\phi^2(x)\,dx$$

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

$$= 0$$

$$\therefore E\left(X^2\Phi(X)\right) = \frac{1}{2}$$

Note :– $\Phi(X) \sim R(0,1)$. $E[\Phi(X)] = \frac{1}{2}$ & $V[\Phi(X)] = \frac{1}{12}$.

$$Cov(X, \Phi(X)) = E[X\Phi(X)] - E(X)\,E(\Phi(X)) = \frac{1}{2\sqrt{\pi}} .$$

Correlation coefficient between X & $\Phi(X)$ is, —

$$\rho_{X,\Phi(X)} = \frac{cov(X, \Phi(X))}{\sqrt{V(X)\,V(\Phi(X))}} = \frac{\frac{1}{2\sqrt{\pi}}}{1 \times \frac{1}{\sqrt{12}}} = \sqrt{\frac{3}{\pi}}$$

$$Cov(X^2, \Phi(X)) = E\left(X^2\Phi(X)\right) - E(X^2)\,E\left(\Phi(X)\right) = \frac{1}{2} - \frac{1}{2} = 0$$

☆ **Problem 5.** Let $Z \sim N(0,1)$ & $G(z) = P[0 \le Z \le z]$ then show that:-

$$\left(\Phi(z)\right)\left(1-\Phi(z)\right) = \frac{1}{4} - G^2(z).$$

Soln. →

$$\begin{aligned} G(z) &= P[0 \le Z \le z] \\ &= \Phi(z) - \Phi(0) \\ &= \Phi(z) - \frac{1}{2} \end{aligned}$$

$$\therefore \Phi(z) = G(z) + \frac{1}{2}$$

$$\begin{aligned} \therefore \quad \Phi(z)\left[1-\Phi(z)\right] &= \left[G(z)+\frac{1}{2}\right]\left[1/2 - G(z)\right] \\ &= \frac{1}{4} - G^2(z). \end{aligned}$$

☆ **Problem 6.** If $X \sim N(\mu,1)$, $\mu > 0$ then show that —

$E\left[\dfrac{1-\Phi(x)}{\phi(x)}\right] = \dfrac{1}{\mu}$, where Φ, ϕ, respectively, denotes the cdf and pdf of $N(0,1)$ variables.

Soln. →

$$\begin{aligned} &E\left[\frac{1-\Phi(x)}{\phi(x)}\right] \\ &= \int_{-\infty}^{\infty} \frac{1-\Phi(x)}{\phi(x)} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x-\mu)^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1-\Phi(x)}{\frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}} \cdot e^{-\frac{1}{2}(x-\mu)^2} \frac{dx}{\sqrt{2\pi}} \\ &= \int_{-\infty}^{\infty} \left[1-\Phi(x)\right] \cdot e^{-\frac{1}{2}(\mu^2 - 2\mu x)} dx \\ &= e^{-\frac{1}{2}\mu^2}\left[\left\{\left(1-\Phi(x)\right)\frac{e^{\mu x}}{\mu}\bigg|_{-\infty}^{\infty}\right\} - \frac{1}{\mu}\int_{-\infty}^{\infty} e^{\mu x}\left(-\phi(x)\right)dx\right] \\ &= e^{-\frac{1}{2}\mu^2}\left[0 + \frac{1}{\mu}\int_{-\infty}^{\infty} e^{\mu x}\phi(x)\,dx\right] \\ &= e^{-\frac{1}{2}\mu^2} \cdot \frac{1}{\mu} M_{SNV}(\mu) \end{aligned}$$

[where, $M_{SNV}(\mu)$ = Moment generating function of standard normal variate]

$$\begin{aligned} &= e^{-\frac{1}{2}\mu^2} \cdot \frac{1}{\mu} \cdot e^{\frac{1}{2}\mu^2} \\ &= \frac{1}{\mu}. \end{aligned}$$

Note:-

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}$$

$$\phi'(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \cdot (-x) = -x\,\phi(x)$$

$$\int \phi'(x)\,dx = -\int x\,\phi(x)\,dx$$

$$\therefore -\phi(x) + C = \int x\,\phi(x)\,dx.$$

☆ Problem 7. If $X \sim N(0, \sigma^2)$ then show that, —

$$P[X > x + \frac{c}{x} / X > x] < e^{-c/\sigma^2}.$$

Soln. → $P[X > x + \frac{c}{x}]$

$$= \int_{x+\frac{c}{x}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-t^2/2\sigma^2} dt \quad \left[\text{ let us transform } z = t - \frac{c}{x}\right]$$

$$= \int_{x}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(z+c/x)^2}{2\sigma^2}} dz$$

$$< \int_{x}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(z^2+2c)} dz$$

$$= e^{-c/\sigma^2} \int_{x}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-z^2/2\sigma^2} dz$$

Now, $\left(z + \frac{c}{x}\right)^2 = z^2 + 2 \cdot c \cdot \frac{z}{x} + \frac{c^2}{x^2}$

$$> z^2 + \frac{2cz}{x}$$

$$> z^2 + 2c \quad [\because z > x]$$

$$\therefore -\left(z + \frac{c}{x}\right)^2 < -(z^2+2c)$$

$$\Rightarrow e^{-\left(z+\frac{c}{x}\right)^2} < e^{-(z^2+2c)}$$

$$\therefore \frac{P[X > x + \frac{c}{x}]}{P[X > x]} < e^{-c/\sigma^2}$$

$$\therefore P[X > x + \frac{c}{x} / X > x] < e^{-c/\sigma^2}.$$

☒ QUARTILE DEVIATION OF NORMAL DISTRIBUTION :—

Let $X \sim N(\mu, \sigma^2)$

$$P[X \le \xi_{1/2}] = \frac{1}{2}$$

$$P[X \le \xi_{1/4}] = \frac{1}{4}$$

$$P[X \le \xi_{3/4}] = \frac{3}{4}$$

Now, $P[X \le \xi_p] = p$

$$\therefore P\left[\frac{X-\mu}{\sigma} \le \frac{\xi_p - \mu}{\sigma}\right] = p$$

$$\Rightarrow \Phi\left(\frac{\xi_p - \mu}{\sigma}\right) = p$$

$$\therefore \xi_p = \mu + \sigma\Phi^{-1}(p)$$

$$\therefore \xi_{1/2} = \mu \Rightarrow \text{median} = \mu.$$

$$\xi_{1/4} = \mu + \sigma\Phi^{-1}\left(\frac{1}{4}\right)$$

$$\xi_{3/4} = \mu + \sigma\Phi^{-1}\left(\frac{3}{4}\right)$$

Note, $\Phi(x) = \frac{1}{4} \Rightarrow 1 - \Phi(x) = \frac{3}{4}$

$$\Rightarrow \Phi(-x) = \frac{3}{4}$$

$$\Rightarrow -x = \Phi^{-1}\left(\frac{3}{4}\right)$$

$$\Rightarrow x = \Phi^{-1}\left(\frac{1}{4}\right)$$

$$\therefore \xi_{1/4} = \mu + \Phi^{-1}\left(\frac{1}{4}\right)\sigma$$

$$= \mu - \sigma\Phi^{-1}\left(\frac{3}{4}\right)$$

and $\xi_{3/4} = \mu + \sigma\Phi^{-1}\left(\frac{3}{4}\right)$.

$$\therefore Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \sigma\Phi^{-1}\left(\frac{3}{4}\right).$$

☆ **Problem 8.** Show that for $x>0$,

$$\left(\frac{1}{x}-\frac{1}{x^3}\right)\phi(x) < (1-\Phi(x)) < \frac{\phi(x)}{x}$$

[WBSU'11]

Soln. →

$$1-\Phi(x) = 1-\int_{-\infty}^{x}\phi(t)dt$$

$$= \int_{x}^{\infty}\phi(t)dt = \int_{x}^{\infty} t\cdot\frac{1}{t}\phi(t)dt$$

$$= \left[-\frac{1}{t}\phi(t)\right]_x^{\infty} + \int_x^{\infty}\frac{-1}{t^2}\cdot\phi(t)dt$$

$$= \frac{\phi(x)}{x} - \int_x^{\infty}\frac{\phi(t)}{t^2}dt$$

$$< \frac{\phi(x)}{x} \quad \left[\because \int_x^{\infty}\frac{\phi(t)}{t^2}dt > 0\right]$$

Now,

$$\int_x^{\infty}\phi(t)dt = \frac{\phi(x)}{x} - \int_x^{\infty}\frac{1}{t^3}\cdot t\,\phi(t)dt$$

$$= \frac{\phi(x)}{x} - \left[\left(-\frac{1}{t^3}\phi(t)\right)_x^{\infty} + \int_x^{\infty}\phi(t)\cdot\frac{(-3)}{t^4}dt\right]$$

$$= \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + \int_x^{\infty}\frac{3}{t^4}\phi(t)dt$$

$$> \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} \quad \left[\because \int_x^{\infty}\frac{3}{t^4}\phi(t)dt > 0\right]$$

$$\therefore \left(\frac{1}{x}-\frac{1}{x^3}\right)\phi(x) < 1-\Phi(x) < \frac{\phi(x)}{x}$$

$$\Rightarrow \left(1-\frac{1}{x^2}\right) < \frac{x(1-\Phi(x))}{\phi(x)} < 1$$

$$\therefore \frac{x(1-\Phi(x))}{\phi(x)} \to 1 \text{ as } x\to\infty \quad [\text{By Squeeze's theorem}]$$

☆ **Problem 9.** For $x>0$ such that $\left|\frac{x\Phi(-x)}{\phi(x)} - 1\right| < \frac{1}{x^2}$ [2002]

Soln. →

$$1-\frac{1}{x^2} < \frac{x[1-\Phi(x)]}{\phi(x)} < 1$$

$$\Rightarrow -\frac{1}{x^2} < \frac{x[\Phi(-x)]}{\phi(x)} - 1 < 0 < \frac{1}{x^2} \quad [\text{as } x>0]$$

$$\therefore \left|\frac{x\Phi(-x)}{\phi(x)} - 1\right| < \frac{1}{x^2}$$

☆ **Problem 10.** Show that $\longrightarrow \lim_{x\to\infty} \frac{1-\Phi(x+\frac{a}{x})}{1-\Phi(x)} = e^{-a}$.

ANS: →
$$\lim_{x\to\infty} \frac{1-\Phi(x+\frac{a}{x})}{1-\Phi(x)}$$
$$= \lim_{x\to\infty} \frac{\cancel{f}\phi(x+\frac{a}{x})\left[1-\frac{a}{x^2}\right]}{\cancel{f}\phi(x)}$$
$$= \lim_{x\to\infty} \frac{e^{-\frac{1}{2}(x+\frac{a}{x})^2}\cdot\left[1-\frac{a}{x^2}\right]}{e^{-\frac{1}{2}x^2}}$$
$$= \lim_{x\to\infty} e^{-\frac{1}{2}\left(\frac{a^2}{x^2}+2a\right)}\left[1-\frac{a}{x^2}\right]$$
$$= e^{-a}.$$

Result: – (Alternative way to prove Problem 8.)
$$1-\frac{1}{x^2} \le \frac{x(1-\Phi(x))}{\phi(x)} \le 1, \quad x>0$$

Proof: –
$$1-\Phi(x) = \int_x^\infty \phi(u)\,du \quad, x>0$$
$$\le \frac{1}{x}\int_x^\infty u\,\phi(u)\,du$$
$$= \frac{1}{x}\left[-\phi(u)\right]_x^\infty$$
$$= \frac{\phi(x)}{x}.$$

$$\therefore \frac{x[1-\Phi(x)]}{\phi(x)} \le 1 \quad \ldots\ldots\ldots (1)$$

$$1-\Phi(x) = \int_x^\infty \frac{1}{u}\cdot u\,\phi(u)\,du = \left[\frac{1}{u}(-\phi(u))\right]_x^\infty - \int_x^\infty \frac{1}{u^2}\phi(u)\,du$$
$$= \frac{\phi(x)}{x} - \int_x^\infty \frac{1}{u^3}\cdot u\,\phi(u)\,du$$
$$= \frac{\phi(x)}{x} - \left[-\frac{1}{u^3}\phi(u)\right]_x^\infty + 3\int_x^\infty \frac{\phi(u)}{u^4}\,du$$
$$= \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + 3\int_x^\infty \frac{\phi(u)}{u^4}\,du$$
$$\ge \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3}$$
$$\therefore \frac{x(1-\Phi(x))}{\phi(x)} \ge 1-\frac{1}{x^2} \quad \ldots\ldots (2)$$

Combining (1) & (2), we get, — $1-\frac{1}{x^2} \le \frac{x(1-\Phi(x))}{\phi(x)} \le 1, \; x>0$

$\lim_{x\to\infty}\left(1-\frac{1}{x^2}\right) = 1$, hence by squeeze theorem, $\lim_{x\to\infty} \frac{x(1-\Phi(x))}{\phi(x)} = 1$.

▨ EXPANSION :—

$$1-\Phi(x) \quad , x>0$$

$$= \frac{\varphi(x)}{x} - \frac{\varphi(x)}{x^3} + 3\int_x^\infty \frac{\varphi(u)}{u^4}\,du$$

$$= \frac{\varphi(x)}{x} - \frac{\varphi(x)}{x^3} + 3\int_x^\infty \frac{u\,\varphi(u)}{u^5}\,du$$

$$= \frac{\varphi(x)}{x} - \frac{\varphi(x)}{x^3} + 3\left[-\frac{\varphi(u)}{u^5}\right]_x^\infty - 15\int_x^\infty \frac{\varphi(u)}{u^6}\,du$$

$$= \frac{\varphi(x)}{x} - \frac{1.\varphi(x)}{x^3} + \frac{1.3.\varphi(x)}{x^5} - 1.3.5.\int_x^\infty \frac{\varphi(u)}{u^6}\,du$$

$$= \frac{\varphi(x)}{x} - \frac{1.\varphi(x)}{x^3} + 1.3.\frac{\varphi(x)}{x^5} - 1.3.5.\frac{\varphi(x)}{x^7} + \cdots$$

$$\cdots + (-1)^{n-1}\,1.3.5\ldots(2n-3).\frac{\varphi(x)}{x^{2n-1}}$$

$$+ (-1)^n\,1.3.5.\ldots(2n-1)\int_x^\infty \frac{\varphi(u)}{u^{2n}}\,du \quad \text{------ Series ①}$$

let the last term is $R_n(x)$.

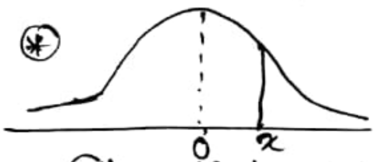
$R_n(x)$ = remainder following n terms

$$|R_n(x)| = 1.3.5.\ldots(2n-1)\int_x^\infty \frac{\varphi(u)}{u^{2n}}\,du$$

$$< 1.3.5.\ldots(2n-1)\,\varphi(x)\int_x^\infty \frac{du}{u^{2n}} \quad \text{------} (*)$$

$$= 1.3.5.\ldots(2n-3)\cancel{(2n-1)}.\varphi(x).\frac{1}{\cancel{(2n-1)}\,x^{2n-1}}.$$

$$= |\text{Previous term}|.$$

(*) 

[If x lies in the right tail of the distn. (significantly away from the origin) then we adopt Series ①]

Given that $x>0$, $\varphi(u) < \varphi(x)\ \forall\ u>x$

$$\therefore \lim_{x\to\infty} \frac{x(1-\Phi(x))}{\varphi(x)}\left\{1-\frac{1}{x^2}+\frac{1.3}{x^4}-\cdots\pm\frac{1.3.5\ldots 2n-1}{x^{2n}}\right\}^{-1} = 1$$

3) An alternative operation of $1-\Phi(x)$:—

$$1-\Phi(x) = \frac{1}{2} - \int_0^x \varphi(u)\,du = \frac{1}{2} - \int_0^x \frac{1}{\sqrt{2\pi}}\left(\sum_{n=0}^\infty \left(-\frac{u^2}{2}\right)^n \Big/ n!\right)du$$

$$= \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\sum_{n=0}^\infty \frac{(-1)^n}{2^n.n!}\int_0^x u^{2n}\,du$$

$$= \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\sum_{n=0}^\infty \frac{(-1)^n}{2^n.n!}\left[\frac{x^{2n+1}}{2n+1}\right] \quad \text{------ Series ②}$$

On the other hand we adopt series ② if x falls close to the origin.

TRUNCATED NORMAL DISTRIBUTION:—

Suppose $X \sim N(\mu, \sigma^2)$

Left tail truncation:— Suppose we are to study the distribution of $X / X > K$, the distribution function of $X/X > K$ is

$$P[X \le x / X > K] = \begin{cases} \dfrac{P(K < X \le x)}{P(X > K)} & \text{if } x > K \\ 0 & \text{if } x \le K \end{cases}$$

$$= \begin{cases} \dfrac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{K-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{K-\mu}{\sigma}\right)} & \text{if } x > K \\ 0 & \text{ow} \end{cases}$$

p.d.f. of $X / X > K$ will be,—

$$g(x) = \begin{cases} \dfrac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{K-\mu}{\sigma}\right)} & \text{if } x > K \\ 0 & \text{ow} \end{cases}$$

$$= \frac{1}{\sigma c}\phi\left(\frac{x-\mu}{\sigma}\right) I_x(K, \infty)$$

when, $c = 1 - \Phi\left(\frac{K-\mu}{\sigma}\right)$

Note that, $g(\mu - x) = g(\mu + x)$.

Since ϕ is an even function but the distribution will not be symmetric as the fact does not hold good for any real x.

NOTE: For real life situations, if we are to fit a truncated normal distribution with left tail truncation then it would be natural that K will be less than the mode of the distribution, i.e. μ.

In that case the mode of the truncated distribution also remains the same.

In such a case, it is worthless to study the shape of the distribution (skewness, kurtosis), we only work out mean & variance of the distribution in order to have an idea about the spread and location of the distribution.

Calculation of mean & variance :

$E[(X-\mu)/X>K]$

$$= \int_K^\infty (x-\mu)\cdot\frac{1}{\sigma c}\,\varphi\left(\frac{x-\mu}{\sigma}\right)dx$$

$$= \frac{\sigma}{c}\int_{\frac{K-\mu}{\sigma}}^{\infty} z\varphi(z)\,dz \qquad \left[z=\frac{x-\mu}{\sigma}\right]$$

$$= \frac{\sigma}{c}\left(-\varphi(z)\right)\Big|_{\frac{K-\mu}{\sigma}}^{\infty}$$

$$= \frac{\sigma\varphi\left(\frac{K-\mu}{\sigma}\right)}{1-\Phi\left(\frac{K-\mu}{\sigma}\right)}$$

$$\therefore E[X/X>K] = \mu + \frac{\sigma\varphi\left(\frac{K-\mu}{\sigma}\right)}{1-\Phi\left(\frac{K-\mu}{\sigma}\right)}$$

$$V(X/X>K) = E\left[(X-E(X/X>K))^2/X>K\right]$$

$$= E\left[(X-\mu)^2/X>K\right] - E^2\left[X-\mu/X>K\right]$$

Note that,

$E\left[(X-\mu)^2/X>K\right]$

$$= \int_K^\infty \frac{(x-\mu)^2}{\sigma c}\,\varphi\left(\frac{x-\mu}{\sigma}\right)dx$$

$$= \frac{\sigma^2}{c}\int_{\frac{K-\mu}{\sigma}}^{\infty} z^2\varphi(z)\,dz \qquad \text{let, } \frac{x-\mu}{\sigma}=z,\quad dx=\sigma dz$$

$$= \frac{\sigma^2}{c}\left[z(-\varphi(z))\Big|_{\frac{K-\mu}{\sigma}}^{\infty} + \int_{\frac{K-\mu}{\sigma}}^{\infty}\varphi(z)\,dz\right]$$

$$= \frac{\sigma^2}{c}\left[+\frac{K-\mu}{\sigma}\varphi\left(\frac{K-\mu}{\sigma}\right) + \left[\Phi(z)\right]_{\frac{K-\mu}{\sigma}}^{\infty}\right]$$

$$= \frac{\sigma^2}{c}\left[\frac{K-\mu}{\sigma}\varphi\left(\frac{K-\mu}{\sigma}\right) + 1-\Phi\left(\frac{K-\mu}{\sigma}\right)\right]$$

$$= \frac{\sigma^2}{c}\left[\frac{K-\mu}{\sigma}\varphi\left(\frac{K-\mu}{\sigma}\right) + c\right]$$

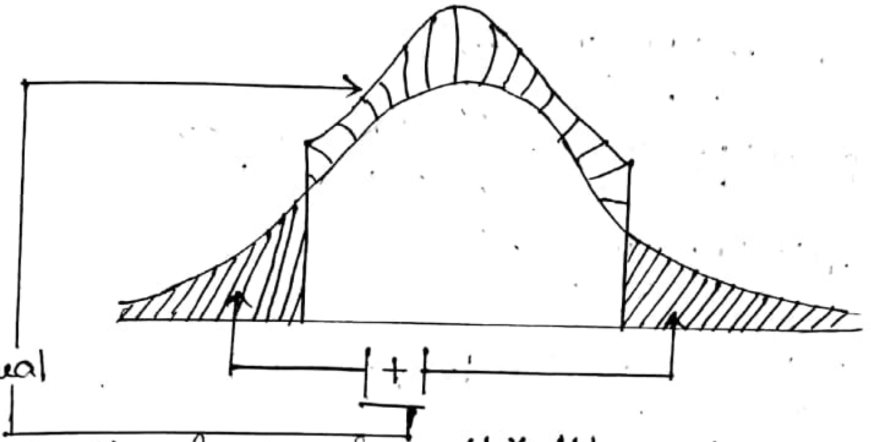
$$= \sigma^2\left[1+\frac{\frac{K-\mu}{\sigma}\varphi\left(\frac{K-\mu}{\sigma}\right)}{1-\Phi\left(\frac{K-\mu}{\sigma}\right)}\right]$$

$$\therefore V(X/X>K) = \sigma^2\left[\left\{1+\frac{\frac{K-\mu}{\sigma}\varphi\left(\frac{K-\mu}{\sigma}\right)}{1-\Phi\left(\frac{K-\mu}{\sigma}\right)}\right\} - \left\{\frac{\varphi\left(\frac{K-\mu}{\sigma}\right)}{1-\Phi\left(\frac{K-\mu}{\sigma}\right)}\right\}^2\right]$$

Symmetric Truncation:— $X \sim N(\mu, \sigma^2)$

Let us consider the distribution of $\left[X \,/\, \left|\frac{X-\mu}{\sigma}\right| < \lambda\right]$ for $\lambda > 0$

i.e. the distribution of $\left[X \,/\, \mu - \lambda\sigma < X < \mu + \lambda\sigma\right]$



The distribution function of $X / \left|\frac{X-\mu}{\sigma}\right| < \lambda$ is

$$G(x) = P\left[X \le x \,/\, \left|\frac{X-\mu}{\sigma}\right| < \lambda\right]$$

$$= \begin{cases} 0 & \text{if } x < \mu - \lambda\sigma \\ \dfrac{P(\mu - \lambda\sigma \le X \le x)}{P(\mu - \lambda\sigma \le X \le \mu + \lambda\sigma)} & \text{if } \mu - \lambda\sigma \le x \le \mu + \lambda\sigma \\ 1 & \text{if } x > \mu + \sigma\lambda. \end{cases}$$

$$= \begin{cases} 0 & \text{if } \frac{x-\mu}{\sigma} < -\lambda \\ \dfrac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi(-\lambda)}{\Phi(\lambda) - \Phi(-\lambda)} & \text{if } \left|\frac{x-\mu}{\sigma}\right| \le \lambda \\ 1 & \text{if } \left(\frac{x-\mu}{\sigma}\right) > \lambda \end{cases}$$

Pdf of $X / \left|\frac{X-\mu}{\sigma}\right| < \lambda$ is,

$$g(x) = \begin{cases} \dfrac{\frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)}{2\Phi(\lambda) - 1} & \text{if } \left|\frac{x-\mu}{\sigma}\right| \le \lambda \\ 0 & \text{ow} \end{cases}$$

$$= \frac{1}{\sigma c}\varphi\left(\frac{x-\mu}{\sigma}\right) I_x\left[\mu - \lambda\sigma, \mu + \sigma\lambda\right]$$

where $c = 2\Phi(\lambda) - 1$.

Note that, $g(\mu - x) = g(\mu + x)\ \forall x$.

Hence the resulting distribution remains the symmetry about μ.

Thus, median = mean = mode = μ [as the distn. is unimodal & bell-shaped].

NOTE: Mean of the distribution exists. Clearly, we need to study the location of the distribution and the skewness of the distribution. Here we study the spread and kurtosis as a feature of the shape of the distribution. Thus it is enough to obtain μ_1 & μ_2 of the distribution.

Now, $\mu = E\left[X / \left|\frac{X-\mu}{\sigma}\right| < \lambda\right]$

$$E\left[\left(X - E\left(X / \left|\frac{X-\mu}{\sigma}\right| < \lambda\right)\right)^2 / \left|\frac{X-\mu}{\sigma}\right| < \lambda\right]$$

$$= E\left[(X-\mu)^2 / \left|\frac{X-\mu}{\sigma}\right| < \lambda\right]$$

$$= \int_{\mu-\lambda\sigma}^{\mu+\lambda\sigma} (x-\mu)^2 \frac{1}{\sigma c} \varphi\left(\frac{x-\mu}{\sigma}\right) dx, \text{ where } c = 2\Phi(\lambda) - 1.$$

$$= \frac{\sigma^2}{c}\int_{-\lambda}^{\lambda} z^2 \varphi(z)\,dz = \frac{2\sigma^2}{c}\int_0^{\lambda} z^2 \varphi(z)\,dz$$

$$= \frac{2\sigma^2}{c}\left[-z\varphi(z)\Big|_0^{\lambda} + \Phi(z)\Big|_0^{\lambda}\right]$$

$$= \frac{2\sigma^2}{c}\left[-\lambda\varphi(\lambda) + \Phi(\lambda) - \Phi(0)\right]$$

$$= \frac{2\sigma^2}{c}\left[\Phi(\lambda) - \frac{1}{2} - \lambda\varphi(\lambda)\right]$$

$$= \frac{2\sigma^2}{c}\left[\frac{c}{2} - \lambda\varphi(\lambda)\right]$$

$$= \sigma^2\left[1 - \frac{\lambda\varphi(\lambda)}{c \cdot \frac{1}{2}}\right]$$

$$= \sigma^2\left[1 - \frac{\lambda\varphi(\lambda)}{\Phi(\lambda) - \Phi(0)}\right]$$

$$= \sigma^2\left[1 - \frac{\lambda\varphi(\lambda)}{\int_0^{\lambda}\varphi(u)\,du}\right]$$

$$= \sigma^2\left[1 - \frac{\lambda e^{-\lambda^2/2}}{I(\lambda)}\right], \text{ where } I(\lambda) = \int_0^{\lambda} e^{-u^2/2}$$

Scanned by CamScanner

Calculation of μ_4 & β_2:—

$$\mu_4 = \int_{\mu-\lambda\sigma}^{\mu+\lambda\sigma} (x-\mu)^4 \cdot \frac{1}{\sigma c} \phi\left(\frac{x-\mu}{\sigma}\right) dx$$

$$= \frac{\sigma^4}{c} \int_{-\lambda}^{\lambda} z^4 \phi(z) dz \qquad \left[\frac{x-\mu}{\sigma} = z\right]$$

$$= \frac{2\sigma^4}{c} \int_0^{\lambda} z^4 \phi(z) dz$$

$$= \frac{2\sigma^4}{c} \left[- z^3 \phi(z) \Big|_0^{\lambda} + 3 \int_0^{\lambda} z^2 \phi(z) dz\right]$$

$$= \frac{2\sigma^4}{c} \left[-\lambda^3 \phi(\lambda) + 3\left[- z\phi(z) \Big|_0^{\lambda} + \int_0^{\lambda} \phi(z) dz\right]\right]$$

$$= \frac{2\sigma^4}{c} \left[-\lambda^3 \phi(\lambda) - 3\lambda\phi(\lambda) + 3\Phi(\lambda) - 3\Phi(0)\right]$$

$$= \frac{2\sigma^4}{c} \left[-\lambda^3 \phi(\lambda) - 3\lambda\phi(\lambda) + 3 \cdot \frac{c}{2}\right]$$

$$= \frac{2\sigma^4}{c} \left[-\lambda\phi(\lambda)(\lambda^2+3) + 3 \cdot \frac{c}{2}\right]$$

$$= \left[3 - \frac{\lambda\phi(\lambda)(\lambda^2+3)}{\int_0^{\lambda} \phi(u) du}\right]$$

$$= \sigma^4 \left[3 - \frac{\lambda e^{-\lambda^2/2}(\lambda^2+3)}{I(\lambda)}\right]$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} > 3$$

$$= \frac{3 - \frac{\lambda(\lambda^2+3) \cdot e^{-\lambda^2/2}}{I(\lambda)}}{\left[1 - \frac{\lambda e^{-\lambda^2/2}}{I(\lambda)}\right]^2} > 3$$

$$\Rightarrow 3 - \frac{\lambda(\lambda^2+3) \cdot e^{-\lambda^2/2}}{I(\lambda)} > 3\left(1 - \frac{\lambda e^{-\lambda^2/2}}{I(\lambda)}\right)^2$$

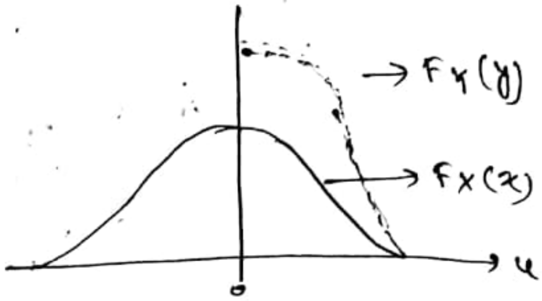
$$\Rightarrow -(\lambda^2+3) > -6 + 3 \cdot \frac{\lambda e^{-\lambda^2/2}}{I(\lambda)}$$

$$\Rightarrow 3 - \lambda^2 > \frac{3\lambda \cdot e^{-\lambda^2/2}}{I(\lambda)} > 0 \qquad \left[\because \frac{\lambda \cdot e^{-\lambda^2/2}}{I(\lambda)} > 0\right]$$

$$\Rightarrow \lambda < \sqrt{3}.$$

☆ **Problem 11.** If $X \sim N(0, \sigma^2)$ find the p.d.f. of $Y = |X|$. Hence or otherwise find $E(|X|)$ and $Var(|X|)$.

Soln. >



$$F_Y(y) = P(Y \le y)$$
$$= P[|X| \le y]$$
$$= P[-y \le X \le y]$$
$$= F_X(y) - F_X(-y)$$
$$= F_X(y) - [1 - F_X(y)]$$
$$= 2F_X(y) - 1$$

$$F_X(y) = \int_{-\infty}^{y} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} dx = \int_{-\infty}^{y} \frac{e^{-z^2/2}\, dz}{\sqrt{2\pi}}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(y) + f_X(-y)$$

$$= \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{-(-y)^2/2\sigma^2} & \text{if } y > 0 \\ 0 & \text{o/w} \end{cases}$$

[Folded Normal distribution]

$$E(|X|) = E(Y) = 2\int_0^{\infty} y \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy$$

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

$$du = \frac{y}{\sigma^2} dy.$$

$$= \sigma\sqrt{\frac{2}{\pi}} \cdot \Gamma 1$$

$$= \sigma\sqrt{\frac{2}{\pi}}$$

$$Var(|X|) = E\{|X|\}^2 - E^2|X|$$
$$= E(X^2) - \sigma^2 \frac{2}{\pi}$$
$$= \sigma^2\left(1 - \frac{2}{\pi}\right).$$

☆ **Problem 12.** If $X \sim N(\mu, \sigma^2)$ then find the median of the distribution.

Soln. →

$f_X(\mu - h) = f_X(\mu + h) \;\forall\, h$ when $X \sim N(\mu, \sigma^2)$

⇒ the p.d.f. is symmetric about μ.

Now, $P[X \le \mu] = \int_{-\infty}^{\mu} f_X(x)\,dx = \int_{-\infty}^{0} f_X(\mu + y)\,dy$ where $x = \mu + y$

$= \int_{-\infty}^{0} f_X(\mu - y)\,dy$, due to symmetry

$= \int_{\infty}^{\mu} f_X(u)(-du)$ where $u = \mu - y$

$= \int_{\mu}^{\infty} f_X(u)\,dy = P(X \geqslant \mu)$

⇒ $x = \mu$ is the median of $X \sim N(\mu, \sigma^2)$.

☆ **Problem 13.** If $X \sim N(\mu, 1)$, $\mu > 0$. find the mean of

$$Y = e^{x^2/2} \int_x^{\infty} e^{-t^2/2}\,dt.$$

Soln. →

$$Y = \frac{1 - \Phi(x)}{\phi(x)}$$

$$E(Y) = E\left\{\frac{1 - \Phi(x)}{\phi(x)}\right\}$$

$$= \int_{-\infty}^{\infty} \frac{1 - \Phi(x)}{\phi(x)} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2}}\,dx$$

$$= e^{-\mu^2/2} \int_{-\infty}^{\infty} \frac{1 - \Phi(x)}{\phi(x)}\, e^{-\mu x}\, \phi(x)\,dx$$

$$= e^{-\mu^2/2} \int_{-\infty}^{\infty} \{1 - \Phi(x)\}\, e^{-\mu x}\,dx$$

$$= e^{-\mu^2/2} \left[\{1 - \Phi(x)\} \frac{e^{\mu x}}{\mu}\right]_{-\infty}^{\infty} - e^{-\frac{1}{2}\mu^2} \int_{-\infty}^{\infty} (-\phi(x)) \frac{e^{\mu x}}{\mu}\,dx$$

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

$$= \frac{1}{\mu} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\mu)^2} \cdot \frac{1}{2\sqrt{\pi}}\,dx$$

$$= \frac{1}{\mu}.$$

$\lim_{x \to -\infty} \{1 - \Phi(x)\} e^{\mu x} = 0$ for $\mu > 0$.

Now, $\lim_{x \to \infty} \{1 - \Phi(x)\} e^{\mu x}$

$= \lim_{x \to \infty} \frac{1 - \Phi(x)}{e^{-\mu x}} \left[\frac{0}{0}\right]$

$= \lim_{x \to \infty} \frac{\phi(x)}{e^{-\mu x}}$ [by L'Hospital's Rule]

$= \lim_{x \to \infty} \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{\mu x}}{e^{x^2/2}} = 0.$

Comparison of Cauchy and Normal Distribution :→

1) If $X \sim C(\mu, \sigma)$ then median = mode = μ. But mean does not exist. The distribution of $C(\mu,\sigma)$ is symmetric about μ.
If $Y \sim N(\mu, \sigma^2)$ then mean = median = mode = μ and the distribution is symmetric about μ.

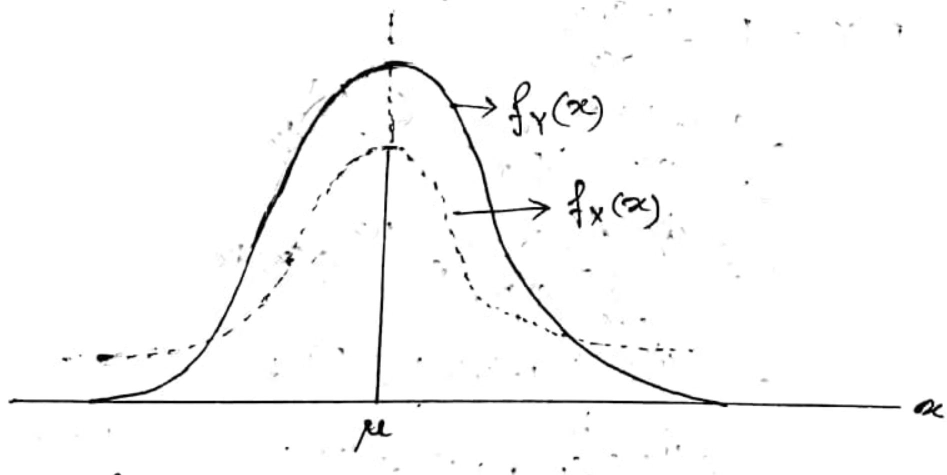
2) $f_X(\mu) = \frac{1}{\pi\sigma}$ and $f_Y(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$ are the maximum ordinates of the PDF s. Here, $f_X(\mu) < f_Y(\mu)$.

3) For large x, $f_Y(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ decreases to zero more rapidly than $f_X(x) = \frac{\sigma}{\pi\{\sigma^2 + (x-\mu)^2\}}$.

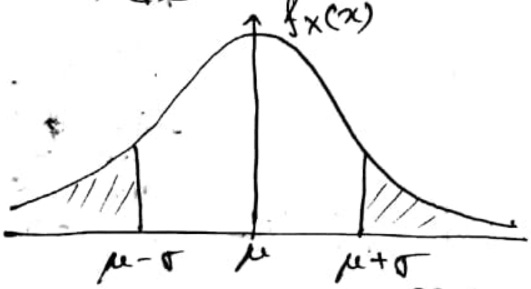
For large x, $f_Y(x) < f_X(x)$

$$\Rightarrow \int_{|u|>x}^{\infty} |u|^r f_Y(u)\,du < \int_{|u|>x}^{\infty} |u|^r f_X(u)\,du$$

Clearly, cauchy distribution has thick tails than normal distribution and consequently mean and μ_r', for $r \geq 1$ does not exist for Cauchy distribution.



4) Point of Inflection :— Let $y = f(x)$ is a twice differentiable function then $\frac{dy}{dx}$ gives the rate of chance of y w.r.t. x. the points for which the rate of chance is minimum or maximum are called point of inflection of $y=f(x)$. Hence, for points of inflection $\frac{d^2y}{dx^2} = 0$ (and $\frac{d^3y}{dx^3} \neq 0$).



For $N(\mu,\sigma^2)$ distribution, the pdf is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; x \in \mathbb{R}$$

Now, $\ln f_X(x) = \text{constant} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$

$$\therefore \frac{f'_X(x)}{f(x)} = -\frac{(x-\mu)}{\sigma^2}$$

$$\Rightarrow f'(x) = -f_X(x)\cdot\left(\frac{x-\mu}{\sigma^2}\right)$$

$$\therefore f_X''(x) = -f_X(x)\frac{1}{\sigma^2} - f_X'(x)\left(\frac{x-\mu}{\sigma^2}\right)$$

$$= -f_X(x)\left\{\frac{1}{\sigma^2} - \frac{(x-\mu)^2}{\sigma^4}\right\}$$

Now, $f_X''(x) = 0 \Rightarrow (x-\mu)^2 = \sigma^2$

$$\Rightarrow x = \mu \pm \sigma.$$

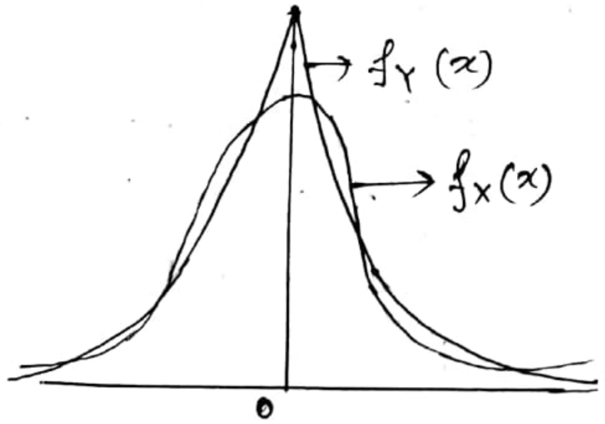
x ——— x

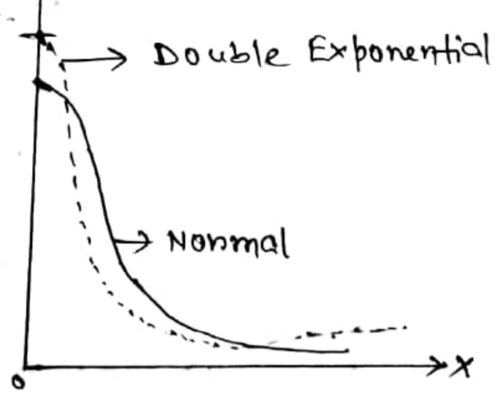
Comparison between Normal & Double Exponential Distribution:-

i) If $X \sim N(0,1)$ and $Y \sim DE(0,1)$, then both X and Y are symmetric about '0' and mean = median = mode = 0.

Here, $f_X(0) = \frac{1}{\sqrt{2\pi}} < \frac{1}{2} = f_Y(0)$

ii) For large $|t|$, $f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ decreases more rapidly than $f_Y(t) = \frac{1}{2}e^{-|t|}$ to zero.





x ——— x

☆ **Problem 14.** Show that —

i) $\left\{\frac{1}{x} - \frac{1}{x^3}\right\}\phi(x) \le 1 - \Phi(x) \le \frac{\phi(x)}{x}$ for $x > 0$

ii) $\lim_{x \to \infty} \frac{x\{1 - \Phi(x)\}}{\phi(x)} = 1$.

Give implication of the limit.

Soln. →

i) For $x > 0$,

$$1 - \Phi(x) = 1 - \int_{-\infty}^{x} \phi(t)\,dt$$

$$= \int_{x}^{\infty} \phi(t)\,dt$$

$$= \int_{x}^{\infty} \frac{1}{t} \cdot t\phi(t)\,dt$$

$$= \left[\frac{1}{t}\left\{-\phi(t)\right\}\right]_{x}^{\infty} - \int_{x}^{\infty}\left(-\frac{1}{t^2}\right)\left(-\phi(t)\right)dt$$

$\left[\because \frac{d}{dt}\phi(t) = \frac{d}{dt}\left(\frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}\right) = \phi(t) \cdot (-t).\right.$

$\left.\text{Hence } \int t\phi(t)\,dt = -\phi(t) + c\right]$

$$= \frac{\phi(x)}{x} - \int_{x}^{\infty} \frac{1}{t^2}\phi(t)\,dt \quad \text{------} (*)$$

$$= \frac{\phi(x)}{x} - \int_{x}^{\infty} \frac{1}{t^3} \cdot t\phi(t)\,dt$$

$$= \frac{\phi(x)}{x} - \left[\frac{1}{t^3} \cdot \left(-\phi(t)\right)\right]_{x}^{\infty} + \int_{x}^{\infty}\left(-\frac{3}{t^4}\right)\left(-\phi(t)\right)dt$$

$$= \frac{\phi(x)}{x} - \frac{\phi(x)}{t^3} + 3\int_{x}^{\infty} \frac{1}{t^4} \cdot \phi(t)\,dt \quad \text{----} (**)$$

From (*) and (**),

$$\left\{\frac{1}{x} - \frac{1}{x^2}\right\}\phi(x) \le 1 - \Phi(x) \le \frac{\phi(x)}{x}$$

ii) $1-\frac{1}{x^2} \le \frac{x\{1-\Phi(x)\}}{\phi(x)} \le 1$, $x>0$

$\therefore \lim_{x\to\infty}\left(1-\frac{1}{x^2}\right) \le \lim_{x\to\infty} \frac{x\{1-\Phi(x)\}}{\phi(x)} \le 1$

$\Rightarrow \lim_{x\to\infty} \frac{x\{1-\Phi(x)\}}{\phi(x)} = 1$ as $\lim_{x\to\infty}\left(1-\frac{1}{x^2}\right)=1$.

Implication:— For large $x(>0)$, we have

$$1-\Phi(x) \simeq \frac{\phi(x)}{x}.$$

For large x, i.e. for $x>3$, the values of $\Phi(x)$ are not, in general, tabulated in Biometrika and in this case, $\{1-\Phi(x)\}$ may be evaluated approximately by the limit.

X —— X

PARETO DISTRIBUTION:→

• Definition:— A RV X is said to have a Pareto distribution if its PDF is given by,

$$f_X(x) = \begin{cases} \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} & \text{, if } x \ge x_0 \\ 0 & \text{, OW} \end{cases}$$

where $x_0>0$, $\alpha>0$.

CDF:→ $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \begin{cases} 0 & \text{if } x<x_0 \\ \int_{x_0}^{x} \frac{\alpha}{x_0}\left(\frac{x_0}{t}\right)^{\alpha+1} dt & \text{if } x\ge x_0 \end{cases}$

$$= \begin{cases} 0 & \text{, if } x<x_0 \\ \alpha x_0^{\alpha} \cdot \int_{x_0}^{x} t^{-\alpha-1}\,dt, \end{cases}$$

$$= \begin{cases} 0 \\ \alpha \cdot x_0^{\alpha} \cdot \left[\frac{t^{-\alpha}}{-\alpha}\right]_{x_0}^{x} \end{cases}$$

$$= \begin{cases} 0 & \text{if } x<x_0 \\ 1-\left(\frac{x_0}{x}\right)^{\alpha} & \text{if } x\ge x_0 \end{cases}$$

- **Definition :→** A continuous RV X is said to have a Pareto distribution if its DF is

$$F_X(x) = \begin{cases} 0 & , \text{ if } x < x_0 \\ 1 - \left(\frac{x_0}{x}\right)^{\alpha} & , \text{ if } x \geq x_0 \end{cases}$$

where $x_0 > 0$, $\alpha > 0$

Moments :—

(a) Existance of Moments : —

$$E|X|^r = \int_{x_0}^{\infty} |x|^r \cdot \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} dx$$

$$= \alpha x_0^{\alpha} \int_{x_0}^{\infty} \frac{1}{x^{\alpha - r + 1}} dx$$

which converges iff $\alpha - r + 1 > 1$

iff $r < \alpha$

Hence $\mu_r' = E(X^r)$ exists iff $r < \alpha$

(b) Moment of order $r\ (< \alpha)$: —

If $r < \alpha$,

$$\mu_r' = E(X^r) = \alpha \cdot x_0^{\alpha} \int_{x_0}^{\infty} \frac{1}{x^{\alpha - r + 1}} dx$$

$$= \alpha x_0^{\alpha} \lim_{u \to \infty} \int_{x_0}^{u} \frac{1}{x^{\alpha - r + 1}} dx$$

$$= \alpha x_0^{\alpha} \lim_{u \to \infty} \left[\frac{x^{-(\alpha - r)}}{-(\alpha - r)} \right]_{x_0}^{u}$$

$$= \alpha x_0^{\alpha} \left\{ \frac{x_0^{-(\alpha - r)} - \lim_{u \to \infty} u^{-(\alpha - r)}}{\alpha - r} \right\}$$

$$= \frac{\alpha}{\alpha - r} \cdot x_0^{r} \quad \text{as } \lim_{u \to \infty} u^{-(\alpha - r)} = 0 \text{ for } r < \alpha.$$

Hence, $E(X) = \mu_1' = \dfrac{\alpha x_0}{\alpha - 1}$ if $\alpha > 1$

$$Var(X) = \mu_2' - \mu_1'^2$$

$$= \frac{\alpha x_0^2}{(\alpha - 1)^2 (\alpha - 2)}, \text{ if } \alpha > 2$$

H.M. : →
$$= \frac{1}{E(X^{-1})} = \frac{1}{\mu'_{-1}} = \frac{1}{\frac{\alpha x_0^{-1}}{\alpha+1}}, \alpha > 1.$$
$$= x_0\left(1+\frac{1}{\alpha}\right)$$

G.M. : →
$$\log G = E(\log_e X)$$
$$= \int_{x_0}^{\infty} \log_e x \cdot f_X(x)\,dx$$

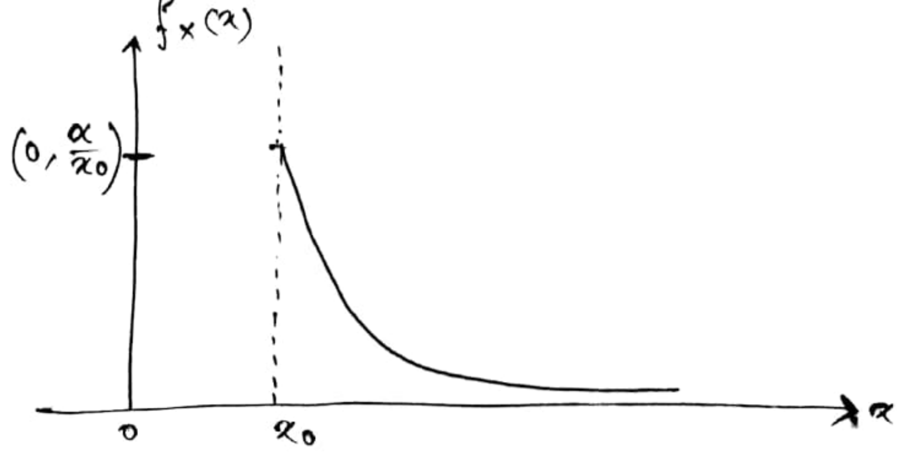
Differentiating $\int_{x_0}^{\infty} \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} dx = 1$, w.r.t. α, we get,
$$\int_{x_0}^{\infty} \left\{ \frac{1}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} + \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} \cdot \log_e\left(\frac{x_0}{x}\right) \right\} dx = 1$$
$$\Rightarrow \frac{1}{\alpha}\int_{x_0}^{\infty} f_X(x)\,dx + \int_{x_0}^{\infty} \log_e\left(\frac{x_0}{x}\right) \cdot f_X(x)\,dx = 0$$
$$\Rightarrow \frac{1}{\alpha}\cdot 1 + E\left\{\log\left(\frac{x_0}{X}\right)\right\} = 0$$
$$\Rightarrow \frac{1}{\alpha} + \log_e x_0 - E(\log_e X) = 0$$
$$\Rightarrow \log G = E(\log_e X) = \log_e\left(x_0 \cdot e^{1/\alpha}\right)$$
$$\Rightarrow G = x_0 \cdot e^{1/\alpha}.$$

Median : →
$$\frac{1}{2} = F_X(\xi_{1/2}) = 1 - \left(\frac{x_0}{\xi_{1/2}}\right)^{\alpha}$$
$$\Rightarrow \xi_{1/2} = x_0 \cdot 2^{1/\alpha}.$$

Mode : → For $x \geq x_0$, $f_X(x) = \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1}$ decreases as x increases.

Hence, $f_X(x)$ is maximum at $x = x_0$,

i.e. the mode of the distribution $= x_0$.



Non existance of MGF :

$$E(e^{tx}) = \int_{x_0}^{\infty} e^{tx} \cdot \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} dx$$

$$= \int_{x_0}^{\infty} \frac{e^{tx}}{x^{\alpha+1}}\, dx \times \frac{\alpha}{x_0^{\alpha}}$$

For $t>0$, $\lim_{x\to\infty} \frac{x^\alpha}{e^{tx}} = 0$, since e^{tx}, $x>0$, increases more rapidly than x^α for large values of x.

For any $\epsilon>0$; $\exists$ a 'b' such that,

$$\frac{x^\alpha}{e^{tx}} < \epsilon \;\; \forall \; x>b.$$

$$\Rightarrow \frac{e^{tx}}{x^{\alpha+1}} > \frac{1}{\epsilon}\cdot\frac{1}{x} \;\; \forall \; x>b$$

For $t>0$,

$$\int_b^{\infty} \frac{e^{tx}}{x^{\alpha+1}} > \frac{1}{\epsilon}\int_b^{\infty} \frac{1}{x}$$

$$= \frac{1}{\epsilon} \lim_{u\to\infty} \int_b^{u} \frac{1}{x}\, dx$$

$$= \frac{1}{\epsilon} \lim_{u\to\infty} [\ln u - \ln b]$$

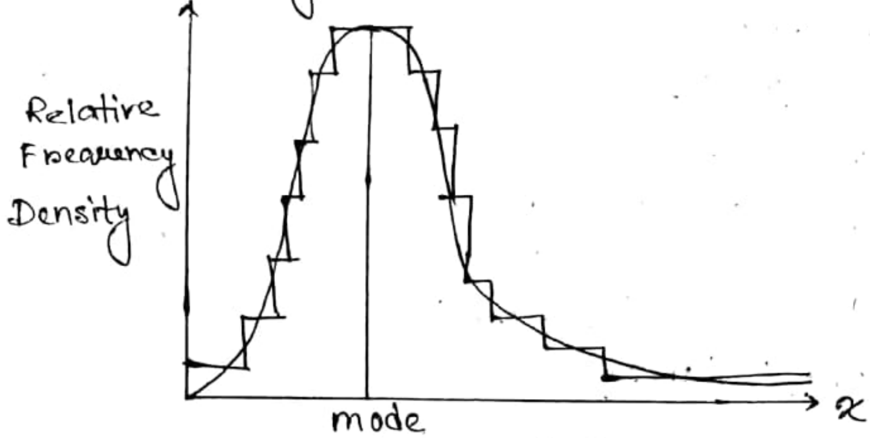
$$= +\infty.$$

For $t>0$, $E[e^{tx}] = +\infty$.

Hence, MGF of Pareto distribution does not exist.

LOGNORMAL DISTRIBUTION :—

In order to graduate (smoothing out the irregularities) an observed income distribution is log-normal and pareto distribution of an adopted. For a typical wealth distribution usually a mixture of the two employed to the hump (mode) of the distribution usually a lognormal density is fitted and beyond it a Pareto density is used.



—: Lognormal Distribution: —

$\therefore$ X is said to have a Lognormal distribution if $\ln X$ is normally distributed.

From the picture it is clear that the lognormal distribution is a positively skewed and bell-shaped distribution.

- **Definition:** — An absolutely continuous random variable X defined over $(0, \infty)$ is said to follow lognormal distribution with parameter μ and σ^2 if its pdf is given by, —

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2}, \quad \begin{array}{l} 0 < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{array}$$

We write, $X \sim \Lambda(\mu, \sigma^2)$.

Note:— $\ln X \sim N(\mu, \sigma^2)$.

$$P[X \le x] = P[\ln X \le \ln x] \quad [\because \ln x \text{ is an increasing function of } x]$$

$$= \begin{cases} \Phi\left(\frac{\ln x - \mu}{\sigma}\right) & \text{if } x > 0 \\ 0 & \text{OW} \end{cases}$$

Probability Density.

$$f(x) = \frac{1}{x\sigma} \varphi\left(\frac{\ln x - \mu}{\sigma}\right) I_x(0, \infty).$$

Result:- $X \sim \Lambda(\mu, \sigma^2)$
$\Leftrightarrow \ln X \sim N(\mu, \sigma^2)$.

Proof:- If part: $\rightarrow$ Let $Y = \ln X \sim N(\mu, \sigma^2)$.

To show, $X \sim \Lambda(\mu, \sigma^2)$.

Distribution function of X is

$$F(x) = P[X \le x]$$
$$= P[Y \le \ln x]$$
$$= G(\ln x), \text{ where } G \text{ denotes the DF of } Y.$$

$\therefore$ The pdf of X is, —

$$f(x) = \frac{d}{dx} G(\ln x) = \frac{1}{x} g(\ln(x))$$
$$= \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} \quad ; 0 < x < \infty$$

$\therefore X \sim \Lambda(\mu, \sigma^2)$.

Only if part: $\rightarrow$ Let $X \sim \Lambda(\mu, \sigma^2)$

To show, $Y = \ln X \sim N(\mu, \sigma^2)$.

The distribution function of Y is, —

$$G(y) = P[Y \le y] = P[\ln X \le y]$$
$$= P[X \le e^y]$$
$$= F(e^y).$$

$\because F$ is distribution function of X.

$\therefore$ The pdf of Y is, —

$$g(y) = \frac{d}{dy} F(e^y) = e^y f(e^y)$$
$$= \cancel{e^y} \cdot \frac{1}{\cancel{e^y} \cdot \sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}, \quad -\infty < y < \infty$$

$$\therefore g(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(y-\mu)^2}, \quad -\infty < y < \infty$$

$\therefore Y \sim N(\mu, \sigma^2)$.

Moments:—

Let $X \sim \Lambda(\mu, \sigma^2)$

Then the rth order raw moment of X is, —

$$\mu_r' = E(X^r) = E(e^{rY}) = M_Y(r) \quad ; Y = \ln X \quad \therefore X = e^Y$$
$$= e^{\mu r + \frac{1}{2}\sigma^2 r^2} ;$$

$$\text{Mean} = e^{\mu} \cdot \delta = e^{\mu + \frac{\sigma^2}{2}} \qquad = e^{\mu r} \cdot \delta^{r^2} \quad [\text{where, } \delta = e^{\sigma^2/2}]$$

(OR) $\mu_r' = \int_0^\infty x^r f(x)\,dx$

$$= \int_0^\infty x^r \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2} \quad ; \quad y=\ln x \quad \therefore dy = \frac{1}{x}dx$$

$$= \int_{-\infty}^{\infty} \frac{e^{ry}}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2} dy$$

$$= M_Y(r)$$

$$= e^{\mu r + \frac{1}{2}\sigma^2 r^2}$$

$$= e^{\mu r} \cdot \delta^{r^2} \quad \left[\text{where } \delta = e^{\sigma^2/2} \right]$$

Putting $r=1$,

$$\mu_1' = e^{\mu+\frac{1}{2}\sigma^2} = E(X).$$

Putting $r=2$, $\mu_2' = \delta^4 e^{2\mu}$.

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

$$= \delta^4 e^{2\mu} - \delta^2 e^{2\mu}$$

$$= e^{2\mu}\delta^2(\delta^2-1).$$

; $\mu_3' = \delta^9 e^{3\mu}$

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

$$= e^{3\mu}(\delta^9 - 3\delta^5 + 2\delta^3)$$

$$= e^{3\mu}(\delta-1)^2(\delta+1)^2\delta^3(\delta^2+2).$$

Quantile:– Let $X \sim \Lambda(\mu,\sigma^2)$ and ξ_p $(0<p<1)$ denotes the pth quantile of X.

Let, η_p denotes the pth quantile of $N(\mu,\sigma^2)$.

Now, $P[X \le \xi_p] = p$

$\Rightarrow P[\ln X \le \ln \xi_p] = p$ [$\because$ $\ln x$ is an increasing function of x]

$\therefore \eta_p = \ln \xi_p$

$\therefore \xi_p = e^{\eta_p}$

$\therefore \xi_{1/2} = e^{\mu}$, $\xi_{1/4} = e^{\mu - .67\sigma}$

& $\xi_{3/4} = e^{\mu + .67\sigma}$

Note:– If $X \sim \Lambda(\mu,\sigma^2)$ then mean > median > mode

$$\Rightarrow e^{\mu+\sigma^2/2} > e^{\mu} > e^{\mu-\sigma^2}.$$

Hence, the lognormal distribution is positively skewed.

Mode of the distribution :—

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} I_x(0,\infty)$$

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}}\left[\frac{1}{x} \cdot e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} \cdot 2 \cdot \left(-\frac{1}{2}\right) \cdot \left(\frac{\ln x-\mu}{\sigma}\right) \cdot \frac{1}{x\sigma} + e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} \cdot \left(-\frac{1}{x^2}\right)\right] = 0$$

$$\Rightarrow \left[-\frac{\ln x-\mu}{\sigma^2} - 1\right] = 0 \qquad [\because x>0]$$

$$\Rightarrow e^{\mu-\sigma^2} = x.$$

Now, $f'(x) = -f(x)\frac{\ln x-\mu}{x\sigma^2} - \frac{1}{x}f(x)$

$$= \frac{1}{x}f(x)\left\{\frac{\ln x-\mu}{\sigma^2}+1\right\}(-1)$$

$$\therefore f''(x) = f(x)\left(\frac{1}{x^2}\right)\left(\frac{\ln x-\mu}{\sigma^2}+1\right) + f(x) \cdot \frac{1}{x^2}\left(\frac{\ln x-\mu}{\sigma^2}+1\right) - \frac{f(x)}{x^2\sigma^2}$$

$$\therefore f''(e^{\mu-\sigma^2}) = f(e^{\mu-\sigma^2})\frac{1}{e^{\mu-\sigma^2}}\left[-\frac{1}{\sigma^2}\right] < 0$$

$\therefore f(x)$ has a maxima at $x = e^{\mu-\sigma^2}$

$\therefore$ Mode $= e^{\mu-\sigma^2}$

Reproductive Property of Lognormal Distribution :—

Suppose X_1 and X_2 are independently distributed lognormal variables, where,

$$X_1 \sim \Lambda(\mu_1, \sigma_1^2)$$
$$X_2 \sim \Lambda(\mu_2, \sigma_2^2)$$

Define, $Y = aX_1^b X_2^c$, $a>0$

a, b, c are non stochastic quantities,

clearly, $\ln Y = \ln a + b\ln X_1 + c\ln X_2$.

where, $\ln X_1$ and $\ln X_2$ are independently distributed normal variates.

where, $\ln X_i \sim N(\mu_i, \sigma_i^2)$, $i=1,2$.

Then the reproductive property of normal distribution, we have,—

$$\ln Y \sim N(\ln a + b\mu_1 + c\mu_2,\ b^2\sigma_1^2 + c^2\sigma_2^2)$$

$$\therefore Y \sim \Lambda(\ln a + b\mu_1 + c\mu_2,\ b^2\sigma_1^2 + c^2\sigma_2^2).$$

Some particular cases:—

i) $X \sim \Lambda(\mu, \sigma^2)$

$aX \sim \Lambda(\ln a + \mu, \sigma^2)$, $a>0$

ii) $X^b \sim \Lambda(b\mu, b^2\sigma^2)$

$\therefore \frac{1}{X} \sim \Lambda(-\mu, \sigma^2)$.

$X \overset{D}{=} \frac{1}{X}$ [if $\mu = 0$]

iii) $X_1X_2 \sim \Lambda(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

$\frac{X_1}{X_2} \sim \Lambda(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2)$

$X_1X_2 \overset{D}{=} \frac{X_1}{X_2}$ [if $\mu_2 = 0$]

☆ Problem 1. If $X \sim \Lambda(\mu, \sigma^2)$ then AM > GM > HM.

Soln. → $AM = E(X)$, $GM = e^{E(\ln X)}$, $HM = \frac{1}{E(\frac{1}{X})}$.

Here, $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} I_x(0,\infty)$.

$\therefore E(X) = e^{\mu+\frac{\sigma^2}{2}}$.

$$E(\ln X) = \int_0^\infty \frac{\ln x}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} dx$$

$$= \int_{-\infty}^\infty \frac{z}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dx \quad , \quad z = \ln x \quad \therefore dz = \frac{1}{x}dx$$

$$= \mu$$

$\therefore GM = e^{\mu}$.

$$E\left(\frac{1}{X}\right) = \int_0^\infty \frac{1}{x} \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} dx$$

$$= \int_{-\infty}^\infty e^{-z} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} dz$$

$$= e^{-\mu+\sigma^2/2}$$

$\therefore HM = e^{\mu-\sigma^2/2}$

$\therefore$ AM > GM > HM.

Mean Deviation about mean:

Let, $X \sim \Lambda(\mu, \sigma^2)$

$$\mu_1' = e^{\mu + \sigma^2/2}$$

$$MD_{\mu_1'}(X) = E|X - \mu_1'|$$

$$= \int_0^\infty |x - \mu_1'| f(x)\,dx$$

$$= 2\int_{\mu_1'}^\infty (x - \mu_1') f(x)\,dx$$

$$= 2\mu_1' \int_{\mu_1'}^\infty \left(\frac{x}{\mu_1'} - 1\right) f(x)\,dx$$

$$= 2\mu_1' \int_{\mu_1'}^\infty \frac{x}{\mu_1'} f(x)\,dx - 2\mu_1' \int_{\mu_1'}^\infty f(x)\,dx$$

$$= 2\mu_1' [I_1 - I_2]$$

$$I_1 = \int_{\mu_1'}^\infty \frac{x}{\mu_1'} f(x)\,dx = \int_{\mu_1'}^\infty \frac{x}{e^{\mu + \frac{\sigma^2}{2}}} \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} dx$$

$$= \int_{\mu_1'}^\infty \frac{e^{\ln x - \mu - \sigma^2/2}}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\{(\ln x)^2 + \mu^2 - 2\mu \ln x\}} dx$$

$$= \int_{\mu_1'}^\infty \frac{e^{-\frac{1}{2\sigma^2}[(\ln x)^2 + \mu^2 + \sigma^4 - 2\mu\ln x - 2\sigma^2 \ln x + 2\mu\sigma^2]}}{x\sigma\sqrt{2\pi}} dx$$

$$= \int_{\mu_1'}^\infty \frac{e^{-\frac{1}{2\sigma^2}[\ln x - \mu - \sigma^2]^2}}{x\sigma\sqrt{2\pi}} \cdot dx$$

$$= P[X' > \mu_1'] \quad \text{where } X' \sim \Lambda(\mu + \sigma^2, \sigma^2)$$

$$= P[\ln X' > \ln \mu_1'] \quad \text{where } \ln X' \sim N(\mu + \sigma^2, \sigma^2)$$

$$= 1 - P\left[\frac{\ln X - \mu - \sigma^2}{\sigma} \le \frac{\ln \mu_1' - \mu - \sigma^2}{\sigma}\right]$$

$$= 1 - \Phi\left[\frac{\ln \mu_1' - \mu - \sigma^2}{\sigma}\right]$$

$$= 1 - \Phi(-\sigma/2)$$

$$= \Phi\left(\frac{\sigma}{2}\right).$$

$$I_2 = \int_{\mu_1'}^{\infty} \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} dx$$

$= P[X > \mu_1']$

$= P[\ln X > \ln \mu_1']$

$= 1 - P[\ln X \le \ln \mu_1']$

$= 1 - \Phi\left(\frac{\ln \mu_1' - \mu}{\sigma}\right)$

$= 1 - \Phi\left(\frac{\mu + \sigma^2/2 - \mu}{\sigma}\right)$

$= 1 - \Phi\left(\frac{\sigma}{2}\right).$

$\therefore MD_{\mu_1'}(X) = 2\mu_1' \left[\Phi\left(\frac{\sigma}{2}\right) - 1 + \Phi(\sigma/2)\right]$

$= 2\mu_1' \left[2\Phi\left(\frac{\sigma}{2}\right) - 1\right]$

☆ **Problem 2.** If $X \sim \Lambda(\mu, \sigma^2)$ then find, $E\left[\frac{X}{E(X)} \ln\left\{\frac{X}{E(X)}\right\}\right]$

Soln. → $E\left[\frac{X}{E(X)} \ln\left\{\frac{X}{E(X)}\right\}\right]$

$= E\left[\frac{X}{E(X)}\left\{\ln X - \ln E(X)\right\}\right]$

$= \frac{E\{X \ln X\}}{E(X)} - \ln\{E(X)\}$

Now, $\frac{1}{E(X)} E(X \ln X) = \frac{1}{E(X)} \int_0^{\infty} x \ln x \cdot \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2} dx$

$$= \int_0^{\infty} \frac{\ln x \cdot e^{(\ln x - \mu - \sigma^2/2) - \frac{1}{2\sigma^2}\{(\ln x)^2 - 2\mu \ln x + \mu^2\}}}{x\sigma\sqrt{2\pi}} dx$$

$$= \int_0^{\infty} \frac{\ln x}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}[\ln x - \mu - \sigma^2]^2} dx$$

$$= \int_{-\infty}^{\infty} \frac{z}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(z - \mu - \sigma^2)^2} dz \qquad \left[\begin{array}{l}\text{let, } \ln x = z \\ \frac{1}{x} dx = dz\end{array}\right]$$

$= E(Z) = \mu + \sigma^2.$

and $\ln\{E(X)\} = \ln\left\{e^{\mu + \sigma^2/2}\right\}$

$= \mu + \sigma^2/2$

Now, $E\left[\frac{X}{E(X)} \ln\left\{\frac{X}{E(X)}\right\}\right] = \mu + \sigma^2 - \mu - \sigma^2/2$

$= \sigma^2/2.$

☆ **Problem 3.** If $X \sim \Lambda(\mu, \sigma^2)$ then μ_r' exists for all $r = 0, 1, \ldots$, but MGF does not exist.

ANS:- If possible suppose $M_X(t)$ exists.

$$M_X(t) = E(e^{tX})$$
$$= E\left[\sum_{r=0}^{\infty} \frac{(tx)^r}{r!}\right]$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \cdot \mu_r'$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \cdot e^{\mu r + \frac{1}{2} r^2 \sigma^2}$$
$$= \sum_{r=0}^{\infty} u_r, \text{ say.}$$

Considering Ratio test to check either the series is convergent or not.

$$\frac{u_{r+1}}{u_r} = \frac{\dfrac{t^{r+1} \cdot e^{\mu(r+1) + \frac{1}{2}(r+1)^2 \sigma^2}}{(r+1)!}}{\dfrac{t^r \cdot e^{\mu r + \frac{1}{2} r^2 \sigma^2}}{r!}} = \frac{t}{r+1} \cdot e^{\mu + \frac{1}{2}(\cancel{r^2\sigma^2} + \sigma^2 + 2r\sigma^2 - \cancel{r^2\sigma^2})}$$

$$= \frac{t}{r+1} \cdot e^{\mu + \frac{1}{2}(\sigma^2 + 2r\sigma^2)} \longrightarrow \infty \text{ as } r \uparrow \infty$$

$\therefore \sum u_r$ is a divergent series.

$\therefore$ it's a contradiction to the fact that MGF exists.

$\therefore$ MGF does not exist.

☆ **Problem 4.** Let X be a RV defined over $(0, \infty)$ with pdf $f_1(x)$ and Y denote another RV on $(0, \infty)$ with p.d.f $f_2(x)$ such that, —

$$E(X^r) = E(Y^r) \quad \forall\, r = 0, 1, 2, \ldots$$

Define a RV Z on $\mathbb{R}$ with pdf. $f(z)$ such that,

$$f(z) = \begin{cases} \frac{1}{2} f_1(z) & \text{if } z > 0 \\ \frac{1}{2} f_2(-z) & \text{if } z < 0 \end{cases}$$

Show that the distribution of Z is not symmetric about '0' but all odd ordered central moment of Z are equal to zero.

ANS:-

$$E(Z^{2r+1}) = \int_0^{\infty} z^{2r+1} \cdot \frac{1}{2} f_1(z)\, dz + \int_{-\infty}^{0} z^{2r+1} \cdot \frac{1}{2} \cdot f_2(-z)\, dz$$
$$= \frac{1}{2} E(X^{2r+1}) + \int_{\infty}^{0} (-z)^{2r+1} \cdot \frac{1}{2} f_2(z)\, dz$$
$$= \frac{1}{2} E(X^{2r+1}) - \frac{1}{2} E(Y^{2r+1})$$
$$= 0 \qquad [\because E(X^r) = E(Y^r)\ \forall\, r = 0, 1, \ldots]$$

But the Distribution of Z is not symmetric about zero, because $f(z) \neq f(-z)$.

Scanned by CamScanner

☆ **Problem 5.** A random variable X has pdf

$$f(x) = \frac{1}{x\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\ln x)^2}, \quad 0 < x < \infty.$$

and a random variable Y has pdf

$$g(y) = \frac{1}{y\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\ln y)^2}(1 + a\sin(2\pi \ln y)) \quad ; \quad |a|<1, \ 0<y<\infty.$$

Show that, — $E(X^n) = E(Y^n) \ \forall \ n = 0, 1, 2, \ldots$

Soln. → $E(Y^n) = \int_0^\infty y^n g(y)\,dy$

$$= \int_0^\infty y^n \cdot \frac{1}{y\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\ln y)^2}(1 + a\sin(2\pi \ln y))\,dy$$

$$= \int_0^\infty y^n f(y)\,dy + \int_0^\infty y^n \cdot \frac{1}{y\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(\ln y)^2} \cdot a\sin(2\pi \ln y)\,dy$$

$$= I_1 + I_2$$

$$I_2 = \int_0^\infty e^{nz} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} \cdot a\sin(2\pi z)\,dz \qquad \left[\text{let, } \ln y = z, \ \frac{1}{y}dy = dz\right]$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z^2 - 2nz + n^2) + \frac{n^2}{2}} \cdot a\sin(2\pi z)\,dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(z-n)^2 + \frac{n^2}{2}} \cdot a\sin(2\pi z)\,dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(t^2 - n^2)} \cdot a\sin 2\pi(t+n)\,dt \qquad \left[\text{let, } z - n = t, \ dz = dt\right]$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(t^2 - n^2)} \cdot a\sin 2\pi t\,dt \qquad [\because n \text{ is an integer and } \sin \text{ is a periodic function}]$$

$= 0$ [since, it is an odd function]

and $I_1 = E(X^n)$ $\therefore E(X^n) = E(Y^n)$

Note:- X and Y have same set of moments.

PARETO DISTRIBUTION: —

Pareto distribution is directly formed from Pareto's law, Pareto's law is named after Vilfredo Pareto, Italian born Swiss professor of economics.

- **Pareto's Law :—**

$$N \propto x^{-a} \quad , a > 0$$

where N = no. of individuals in a community having income atleast x. (in appropriate unit)

The constant 'a' determining the law explicitly is called a Pareto constant.

Suppose, X is a random variable denoting the income obviously in appropriate unit.

Clearly, $P[X \geq x]$ can be regarded as the proportion of individual having income atleast x, it immediately follow from the Pareto's law.

$$P[X \geq x] \propto x^{-a}$$

$$\Rightarrow P[X \geq x] = A x^{-a} \quad [A > 0, \text{ being a suitable constant}]$$

If k denotes the minimum income earned by an individual belonging to the community then,

$$P[X \geq k] = 1$$

$$\Rightarrow A k^{-a} = 1$$

$$\Rightarrow A = k^{a}$$

$$\therefore P[X \geq x] = \left(\frac{k}{x}\right)^{a}$$

Thus F denotes the DF then $F(x) = 1 - \left(\frac{k}{x}\right)^{a}$ if $x \geq k$

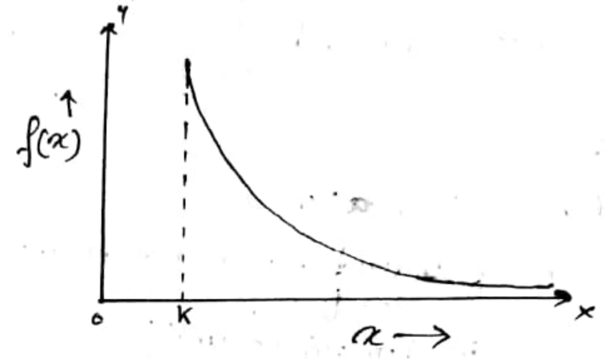
$= 0$ ow

Hence $f(x) = \dfrac{a k^{a}}{x^{a+1}} I_x[k, \infty)$.

- **Definition :—** If X follows Pareto distribution then,

$$f(x) = \begin{cases} \dfrac{a k^{a}}{x^{a+1}} & , k \leq x < \infty \\ 0 & , \text{ow} \end{cases}$$

Features of the distribution :—



Here we see that the pdf is hyperbolic, the distribution is extremely positively skewed (J-shaped), with K = mode of the distribution.

Here we did not study the kurtosis as a feature of the shape of the distribution. The moments of the distn. exists depending on the choice of pareto constant 'a', as some of the integers moments of the distribution do not exist.

Here, MGF does not exist, the Gini's Coefficient of the distribution is, $\frac{1}{(2a-1)}$ with $a>1$.

Moments :—

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

$$= ak^a \int_k^\infty \frac{dx}{x^{(a-r)+1}}$$

$$= ak^a \left[\frac{1}{-(a-r)\, x^{a-r}} \right]_k^\infty$$

$$= \frac{ak^r}{(a-r)}, \quad a>r$$

Mean $= E(X) = \left(\frac{ak}{a-1}\right), \quad a>1$

Variance $= \frac{ak^2}{a-2} - \frac{a^2k^2}{(a-1)^2}, \quad a>2$

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

$$= \frac{ak^2}{(a-2)(a-1)^2}$$

Median :—

$\boxed{\text{C.U.}}$

$$\int_{\xi_{1/2}}^\infty \frac{ak^a}{x^{a+1}}\, dx = \frac{1}{2} \Rightarrow ak^a \left[-\frac{1}{ax^a} \right]_{\xi_{1/2}}^\infty = \frac{1}{2}$$

$$\Rightarrow \cancel{a}k^a \left[\frac{1}{\cancel{a}\xi_{1/2}^a} \right] = \frac{1}{2}$$

$$\Rightarrow \xi_{1/2} = K \cdot 2^{1/a}$$

$\therefore$ Median $= 2^{1/a} \cdot K$.

Mean Deviation About mean: —

$$E|X-\mu| \quad , \quad \mu = \frac{ak}{a-1}, \; a>1$$

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

$$= 2\int_{\mu}^{\infty} \frac{x a k^a}{x^{a+1}}\,dx - 2\mu\int_{\mu}^{\infty} \frac{ak^a}{x^{a+1}}\,dx$$

$$= 2ak^a\left[-\frac{1}{(a-1)x^{a-1}}\Big|_{\mu}^{\infty} + \frac{\mu}{a x^a}\Big|_{\mu}^{\infty}\right]$$

$$= 2ak^a\left[\frac{1}{(a-1)\mu^{a-1}} - \frac{1}{a\mu^{a-1}}\right]$$

$$= \frac{2k^a}{\mu^{a-1}(a-1)}$$

Now,

$$\frac{MD}{S.d} = \frac{\frac{2k^a}{(a-1)} \times \frac{1}{\left(\frac{ak}{a-1}\right)^{a-1}}}{\frac{k}{(a-1)}\sqrt{\frac{a}{a-2}}}, \; a>2$$

$$= \frac{2k^a}{(a-1)} \times \frac{(a-1)^{a-1}}{(ak)^{a-1}} \times \frac{(a-1)\sqrt{a-2}}{\sqrt{a}\times k}$$

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

Mode of the distribution: —

$$f(x) = \frac{ak^a}{x^{a+1}}.$$

Now, differentiating w.r.t. x, we get —

$$f'(x) = -ak^a(a+1)\cdot\frac{1}{x^{a+2}} < 0$$

$\therefore$ $f(x)$ is a decreasing function of x.

$\therefore$ $f(x)$ has its mode at $x=k$,

i.e. $f(x)$ is maximum when x is minimum.

$\therefore$ Mode $= k$.

Scanned by CamScanner

Gini Coefficient of Concentration :—

Gini's mean difference,

$$\Delta = \int_K^\infty \int_K^\infty |x-y| f(x) f(y)\, dx\, dy$$

$$= 2\int_K^\infty F(x)\,[1-F(x)]\, dx$$

$$= 2\int_K^\infty \left(1-\left(\frac{K}{x}\right)^a\right)\left(\frac{K}{x}\right)^a dx$$

$$= 2\int_K^\infty \left(\frac{K}{x}\right)^a dx - 2\int_K^\infty \left(\frac{K}{x}\right)^{2a} dx$$

$$= 2K^a\left[-\frac{1}{(a-1)x^{a-1}}\right]_K^\infty - 2K^{2a}\left[-\frac{1}{(2a-1)x^{2a-1}}\right]_K^\infty$$

$$= 2K^a\left[\frac{1}{(a-1)K^{a-1}}\right] - 2K^{2a}\left[\frac{1}{(2a-1)K^{2a-1}}\right] \quad [\because a>1]$$

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

$$= \frac{2Ka}{(a-1)(2a-1)}$$

Gini Coefficient of Concentration = 2 × Lorenz area

C.U. $\quad \therefore \dfrac{\Delta}{2\mu} = \dfrac{1}{(2a-1)}\,; \quad a>1.$

Mean :–

$$\mu = E(X) = \int_K^\infty x \cdot \frac{aK^a}{x^{a+1}}\, dx$$

$$= aK^a \int_K^\infty x^{-a}\, dx$$

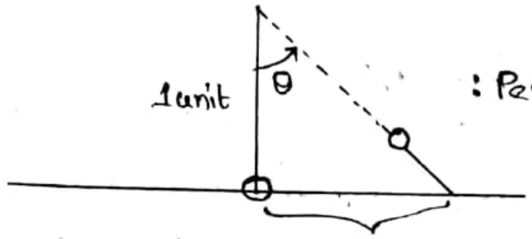
$$= aK^a \cdot \frac{x^{1-a}}{1-a}\Bigg|_K^\infty$$

$$= aK^a \cdot \frac{(-K^{1-a})}{1-a} \quad [\text{provided, } a>1]$$

$$= \frac{aK}{a-1}.$$

CAUCHY DISTRIBUTION :—

Derivation of the p.d.f. : →



: Pendulam spinning at random

Hence, $\theta \sim \text{Uniform}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

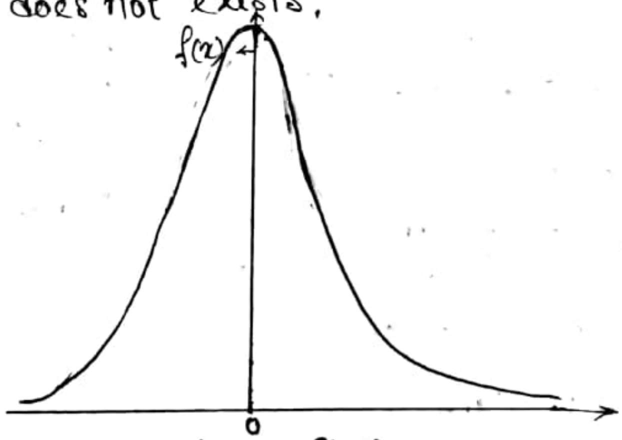
Clearly, $X = \tan\theta$,

$$\therefore P[X \le x] = P[\tan\theta \le x]$$
$$= P[\theta \le \tan^{-1}x]$$
$$= \frac{\tan^{-1}x + \frac{\pi}{2}}{\pi}$$
$$= \frac{1}{2} + \frac{1}{\pi}\tan^{-1}x \quad , \quad x \in \mathbb{R}$$

$\Rightarrow f(x) = \frac{1}{\pi}\cdot\frac{1}{1+x^2}$; $x \in \mathbb{R}$. We write, $X \sim C(0,1)$.

The above is the pdf of Standard Cauchy Distribution.

Clearly, Cauchy distribution is symmetric about zero, i.e. median of the distribution is zero, moreover the distribution is bell-shaped, i.e. mode is also zero but the mean of the distribution does not exists.



— : Standard Cauchy Density function :—

The tails of the cauchy distribution are heavy, i.e. the tail probabilities are significant.

Here it can be easily verified that →

$X \sim$ standard Cauchy $\Leftrightarrow \frac{1}{X} \sim$ Standard Cauchy [typical feature]

It is evident that if $X \sim$ the standard Cauchy distribution then distribution of X and $-X$ will be identical (regular feature)

ONE PARAMETER CAUCHY DISTRIBUTION :—

X is said to be Cauchy variate of median μ if the pdf of x is of the form

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\mu)^2}$$

Clearly, the distribution is symmetric about μ.

We denote, — $X \sim$ Cauchy with median μ.

$$\Leftrightarrow (X-\mu) \sim \text{Standard cauchy}.$$

TWO PARAMETER CAUCHY DISTRIBUTION :—

An absolutely continuous random variable X defined over $(-\infty,\infty)$ is said to follow Cauchy distribution with parameters μ and σ [median $=\mu$, scale $=\sigma$] if its pdf is given by, —

$$f(x) = \frac{\sigma}{\pi} \cdot \frac{1}{[\sigma^2+(x-\mu)^2]} \quad ; \quad -\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0,$$

We denote, $X \sim C(\mu,\sigma)$.

$$\Leftrightarrow \frac{X-\mu}{\sigma} \sim C(0,1).$$

Result:- If $X \sim C(\mu,\sigma)$ then show that $\frac{X-\mu}{\sigma} \sim C(0,1)$.

Soln. →

$X \sim C(\mu,\sigma)$

$$f_X(x) = \frac{\sigma}{\pi[\sigma^2+(x-\mu)^2]}, \quad x \in \mathbb{R}, \; \mu \in \mathbb{R}, \; \sigma \in \mathbb{R}^+.$$

$\left[\text{let, } Z = \frac{X-\mu}{\sigma}\right]$

$$\begin{aligned} F_Z(z) &= P[Z \le z] \\ &= P\left[\frac{X-\mu}{\sigma} \le z\right] \\ &= P[X \le \mu+\sigma z] \\ &= F_X(\mu+\sigma z) \end{aligned}$$

$$\begin{aligned} f_Z(z) = \frac{d}{dz} \cdot F_Z(z) = \frac{d}{dz} F_X(\mu+\sigma z) &= \sigma \cdot f_X(\mu+\sigma z) \\ &= \frac{\sigma^2}{\pi[\sigma^2+(\mu+\sigma z-\mu)^2]} \\ &= \frac{1}{\pi(1+z^2)} \end{aligned}$$

$$\therefore Z = \frac{X-\mu}{\sigma} \sim C(0,1)$$

Reproductive Property of Cauchy Distribution :—

Suppose $X_1, X_2, \ldots, X_n$ are independently distributed cauchy variates where, $X_i \sim C(\mu_i, \sigma_i)$ $\forall\, i = 1, 2, \ldots, n$.

Define, $S_n = \sum_{i=1}^{n} X_i$,

then $S_n \sim \left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i\right)$.

In particular, if X_i's are i.i.d. cauchy variables with median μ and scale σ then

$$S_n \sim C(n\mu, n\sigma).$$

$$\Leftrightarrow \overline{X_n} = \frac{S_n}{n} \sim C(\mu, \sigma).$$

i.e. in case of sampling of cauchy(μ, σ) distribution the sampling distribution of the sample mean will be same as the parent distribution.

Note:– Like normal distribution, Cauchy distribution is also a stable distribution. (Proof by Central Limit theorem).

• Proof of the Reproductive property :→

Result:– If $X \sim C(0,1)$ and $Y \sim C(0,1)$, then $Z = X + Y \sim C(0,2)$.

Proof:– $X \sim C(0,1)$; $f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$; $-\infty < x < \infty$

$Y \sim C(0,1)$; $f(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2}$; $-\infty < y < \infty$

$Z = X + Y$,

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

Now,

$$\frac{1}{(1+x^2)(1+(z-x)^2)} = \frac{1}{z^2(z^2+4)} \left[\frac{2zx}{1+x^2} + \frac{z^2}{1+x^2} + \frac{2z^2 - 2zx}{1+(z-x)^2} + \frac{z^2}{1+(z-x)^2} \right].$$

so that,

$$f_Z(z) = \frac{1}{\pi^2} \cdot \frac{1}{z^2(z^2+4)} \left[z \log \frac{1+x^2}{1+(z-x)^2} + z^2 \tan^{-1} x + z^2 \tan^{-1}(x-z) \right]_{-\infty}^{\infty}$$

$$= \frac{1}{\pi} \cdot \frac{2}{(z^2+2^2)} ; \quad -\infty < z < \infty.$$

$\therefore X + Y \sim C(0, 2)$. [Proved]

Some features of two parameter Cauchy Distribution:-

QUARTILES:-

The distribution function of X is,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} \frac{\sigma\,dt}{\pi[\sigma^2 + (t-\mu)^2]}$$

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

$$= \frac{1}{\pi}\int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{dz}{1+z^2}$$

let, $\frac{t-\mu}{\sigma} = z$

$\frac{1}{\sigma}dt = dz$

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

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

$$F(\xi_p) = p$$

$$\Rightarrow \frac{1}{\pi}\tan^{-1}\left(\frac{\xi_p-\mu}{\sigma}\right) + \frac{1}{2} = p$$

$$\Rightarrow \tan^{-1}\left(\frac{\xi_p-\mu}{\sigma}\right) = \pi\left(p-\frac{1}{2}\right)$$

$$\therefore \xi_p = \mu + \sigma\tan\left\{\pi\left(p-\frac{1}{2}\right)\right\}$$

$$\therefore \xi_{1/2} = \mu;\quad \xi_{1/4} = \mu-\sigma;\quad \xi_{3/4} = \mu+\sigma.$$

$$\therefore \text{Quartile Deviation} = \frac{\cancel{\mu}+\sigma-\cancel{\mu}+\sigma}{2} = \sigma.$$

$$\therefore \text{Median} = \mu.$$

Mode of the Distribution:-

$$f(x) = \frac{\sigma}{\pi}\cdot\frac{1}{[\sigma^2 + (x-\mu)^2]}$$

$$f'(x) = \frac{\sigma}{\pi}\cdot\frac{-2(x-\mu)}{\{\sigma^2+(x-\mu)^2\}^2} = 0$$

$$\therefore x = \mu.$$

$$f''(x) = \frac{\sigma(-2)}{\pi}\cdot\frac{(\sigma^2+(x-\mu)^2)^2 - (x-\mu)^2\,2\,(\sigma^2+(x-\mu)^2).2}{\{\sigma^2+(x-\mu)^2\}^4}$$

$$f''(x) = -\frac{2\sigma}{\pi}\cdot\frac{1}{\sigma^4} < 0$$

$\therefore f(x)$ has its maxima at $x=\mu$.

$$\therefore \text{Mode} = \mu.$$

Maximum ordinate :— at $x = \mu$,

$$f(\mu) = \frac{1}{\pi\sigma}.$$

☆ Problem 1. If $X \sim C(\mu, \sigma)$, then show that the distribution is symmetric about μ.

ANS :—

$$f(\mu + h) = \frac{\sigma}{\pi(\sigma^2 + h^2)}.$$

$$f(\mu - h) = \frac{\sigma}{\pi(\sigma^2 + h^2)}.$$

$\therefore f(\mu+h) = f(\mu-h)$; i.e. the distribution is symmetric about μ.

☆ Problem 2. If $X \sim C(0,1)$ then show that, $E(X^r)$ exists iff $|r| < 1$.

Proof :—

$$E|X|^r = \int_{-\infty}^{\infty} |x^r| \cdot \frac{1}{\pi(1+x^2)} dx$$

$$= \frac{2}{\pi}\int_0^\infty x^r \cdot \frac{dx}{(1+x^2)} \quad \left[\text{assuming } \int_0^\infty \frac{x^r}{1+x^2} dx \text{ converges}\right]$$

$$= \frac{1}{\pi}\int_0^\infty \frac{z^{\left(\frac{1}{2}+\frac{r}{2}\right)-1}}{(1+z)} dz \quad \left[\text{let, } x^2 = z, \ 2x\,dx = dz\right]$$

$$= \frac{1}{\pi}\int_0^\infty \frac{z^{\left(\frac{1}{2}+\frac{r}{2}\right)-1}}{(1+z)^{\left(\frac{1}{2}+\frac{r}{2}\right)+\left(\frac{1}{2}-\frac{r}{2}\right)}} dz$$

For a cauchy distribution fractional moment exists but raw moment does not exist.

$$= \frac{1}{\pi}\cdot \beta\left(\frac{1}{2}+\frac{r}{2}, \frac{1}{2}-\frac{r}{2}\right), \text{ provided, } -1 < r < 1$$

$\frac{r+1}{2} > 0 \Rightarrow r > -1$

$\frac{1-r}{2} > 0 \Rightarrow r < 1$

i.e. $|r| < 1$

$\therefore E|X|^r$ exists or μ_r' exist iff $|r| < 1$.

Note :— If $X \sim C(0,1)$, μ_r' exists for $|r| < 1$.
If $X \sim C(\mu, \sigma^2)$, $E(X-\mu)^r$ exists for $|r| < 1$.

☆ Problem 3. If $X \sim C(0,1)$ then S.T. $\frac{1}{X} \sim C(0,1)$.

Proof :— Let, $Y = \frac{1}{X}$,

$$F_Y(y) = P[Y \le y]$$
$$= P\left[\frac{1}{X} \le y\right]$$
$$= P\left[X \geq \frac{1}{y}\right]$$
$$= 1 - P\left[X < \frac{1}{y}\right]$$
$$= 1 - \int_{-\infty}^{1/y} \frac{dt}{\pi(1+t^2)}$$
$$= 1 - \frac{1}{\pi}\left[\tan^{-1} t\right]_{-\infty}^{1/y}$$
$$= \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left(\frac{1}{y}\right)$$

$$\therefore f_Y(y) = \frac{1}{\pi y^2}\cdot\frac{1}{1+\frac{1}{y^2}} = \frac{1}{\pi(1+y^2)}$$

$\therefore Y \sim C(0,1)$.

☆ **Problem 4.** If $X \sim C(0,1)$ then show that the MGF of X does not exist. [C.U. 2005]

Proof:– $MGF = M_X(t) = E(e^{tX})$

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

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

$$= I_1 + I_2$$

$$I_1 = \int_{-\infty}^{0} \frac{e^{tx}\, dx}{\pi(1+x^2)} < \int_{-\infty}^{0} \frac{dx}{\pi(1+x^2)} \quad [\text{for } t>0,\ e^{tx}<1 \text{ as } x \text{ is negative}]$$

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

$\therefore I_1$ exists if $t>0$.

$$I_1 = \int_{-\infty}^{0} \frac{e^{tx}}{\pi(1+x^2)}\, dx > \int_{-\infty}^{0} \frac{tx}{\pi(1+x^2)} \quad [\text{if } t<0, \text{ then } e^{tx} > tx]$$

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

$$= \frac{t}{2\pi} \lim_{s\to -\infty} \left[\log(1+x^2)\right]_s^0$$

$\therefore I_1$ does not exist for $t<0$.

$$I_2 = \int_{0}^{\infty} e^{tx} \cdot \frac{1}{\pi(1+x^2)}\, dx$$

when $t<0$, $e^{tx}<1$

$\therefore I_2$ exists.

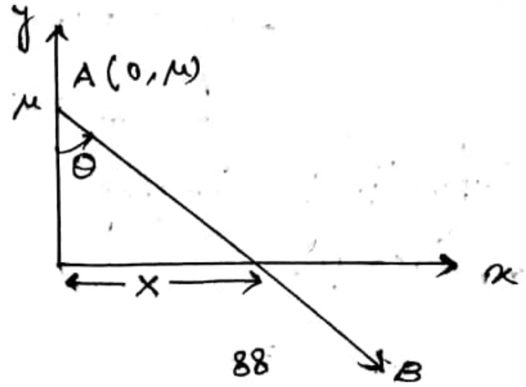
When $t>0$, I_2 does not exist.

Hence, there does not exist any t such that $|t|<h$, for which $E(e^{tX})$ exists.

$\therefore$ MGF does not exist.

☆ **Problem 5.** A straight line AB is free to move through a fixed point A with co-ordinates $(0,\mu)$ and the length of the intersection 'X' it makes with the x-axis is noted. AB makes an angle θ with the y-axis assuming that θ has an uniform distribution between $-\frac{\pi}{2}$ to $\frac{\pi}{2}$. find the distribution of X.

Solution:–



$$\frac{X}{\mu} = \tan\theta$$

$$\Rightarrow X = \mu \tan\theta$$

The distribution function of X is, —

$$F_X(x) = P[X \leq x]$$
$$= P[\mu \tan\theta \leq x]$$
$$= P[\theta \leq \tan^{-1}\frac{x}{\mu}]$$
$$= \int_{-\pi/2}^{\tan^{-1}\frac{x}{\mu}} g(\theta)\,d\theta$$
$$= \int_{-\pi/2}^{\tan^{-1}\frac{x}{\mu}} \frac{1}{\pi}\,d\theta$$
$$= \frac{1}{\pi}\tan^{-1}\frac{x}{\mu} + \frac{1}{2}$$

$\therefore$ Pdf of X is, — $f(x) = \frac{1}{\pi}\cdot\frac{1/\mu}{1+\left(\frac{x}{\mu}\right)^2}$

$$= \frac{1}{\pi}\cdot\frac{\mu}{\mu^2+x^2}.$$

$\therefore X \sim C(0,\mu)$.

☆ Problem 6. If $X \sim C(0,1)$, $Y = \frac{2X}{1+X^2}$, find $E(X)$ & $V(X)$.

ANS :- $f(x) = \frac{1}{\pi(1+x^2)}\, I_x(-\infty,\infty)$.

$$E(Y) = E\left(\frac{2X}{1+X^2}\right) = \int_{-\infty}^{\infty}\frac{2x}{1+x^2}\cdot\frac{1}{\pi(1+x^2)}\,dx$$
$$= 0$$

$$V(Y) = E\left(\frac{2X}{1+X^2}\right)^2 = \int_{-\infty}^{\infty}\frac{4x^2}{\pi(1+x^2)^3}\,dx$$
$$= \frac{4\cdot 2}{\pi}\int_0^{\infty}\frac{x^2}{(1+x^2)^3}\,dx$$
$$= \frac{8}{\pi}\int_0^{\frac{\pi}{2}}\frac{\tan^2 z\cdot\sec^2 z}{\sec^6 z}\,dz \qquad \left[\text{let, } x = \tan z,\ dx = \sec^2 z\,dz\right]$$
$$= \frac{8}{\pi}\int_0^{\pi/2}\sin^2 z\cos^2 z\,dz$$
$$= \frac{8}{\pi}\beta\left(\frac{3}{2},\frac{3}{2}\right)$$
$$= \frac{8}{\pi}\cdot\frac{\Gamma\frac{3}{2}\,\Gamma\frac{3}{2}}{\Gamma 3}$$
$$= 1.$$

☆ Problem 7. $X \sim C(\mu,1)$ then find $E\left(\frac{\partial}{\partial\mu}\log_e f(x)\right)^2$.

Soln. → $f(x) = \frac{1}{\pi\{1+(x-\mu)^2\}} I_x(-\infty,\infty)$.

$\ln f(x) = -\ln\pi - \ln\{1+(x-\mu)^2\}$

$\frac{\partial}{\partial\mu}\ln f(x) = +\frac{2(x-\mu)}{1+(x-\mu)^2} = Y \ldots$ say.

$$\therefore E(Y)^2 = \int_{-\infty}^{\infty} \frac{2^2(x-\mu)^2}{\{1+(x-\mu)^2\}^2} \cdot \frac{1}{\pi(1+(x-\mu)^2)} dx$$

$$= \frac{4}{\pi}\int_{-\infty}^{\infty} \frac{z^2}{(1+z^2)^3} dz \qquad \left[\begin{array}{l}\text{let, } x-\mu = z \\ dx = dz\end{array}\right]$$

$$= \frac{4\cdot 2}{\pi}\int_0^{\infty} \frac{z^2}{(1+z^2)^3} dx$$

$$= 1. \qquad [\text{see the previous problem}]$$

☆ Problem 8. If $X \sim C(\mu,\sigma)$ find the CDF and median of the distn..

Soln. → $$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{x} \frac{\sigma}{\pi\{\sigma^2+(t-\mu)^2\}} dt$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\pi(1+z^2)} dz, \text{ where } z = \frac{t-\mu}{\sigma}$$

$$= \frac{1}{\pi}\lim_{a\to-\infty}\int_a^{\frac{x-\mu}{\sigma}} \frac{1}{1+z^2} dz$$

$$= \frac{1}{\pi}\lim_{a\to-\infty}\left[\tan^{-1} z\right]_a^{\frac{x-\mu}{\sigma}}$$

$$= \frac{1}{\pi}\lim_{a\to-\infty}\left\{\tan^{-1}\left(\frac{x-\mu}{\sigma}\right) - \tan^{-1} a\right\}$$

$$= \frac{1}{\pi}\left\{\tan^{-1}\left(\frac{x-\mu}{\sigma}\right) - \left(-\pi/2\right)\right\}$$

$$= \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}$$

Median:— The median of $X \sim C(\mu,\sigma)$ is given by

$$F_X(x) = \frac{1}{2}$$

$$\Rightarrow \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}$$

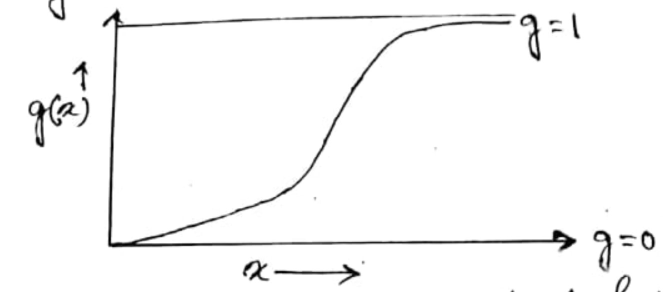
$$\Rightarrow x = \mu$$

Hence $x = \mu$ is the median of $C(\mu,\sigma)$ distribution.

LOGISTIC DISTRIBUTION :—

- **Derivation of the pdf of standard Logistic distribution:—**

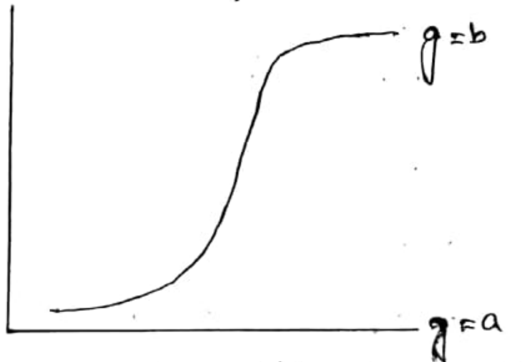
The function $g(x) = \frac{1}{1+e^{-x}}$ is termed as sigmoid function which have asympotes at $x = -\infty$ and $x = +\infty$ and it is skewsymmetric w.r.t. $x = 0$.



In general, a sigmoid curve can be obtained from the differential equation:

$$\frac{dg}{dx} \propto (g-a)(b-g)$$

Where, g is bound to lie in (a, b).



[$g-a \to$ excess over the initial, $b-g \to$ deficiency to the final]

If we choose $a=0$ and $b=1$ then g can be regarded as the distribution function $F(x)$.
The distribution associated with $\frac{dF(x)}{dx} \propto F(x)(1-F(x))$ is known as logistic distribution.

Since the above leads to $\rightarrow$ $\log_e \frac{F(x)}{1-F(x)} = \alpha + \beta x$

where we see that the logit [log of odds is linear in x]

Thus $F(x)$ reduces to,

$$F(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

$$= \frac{1}{1+e^{-(\alpha+\beta x)}}$$, the distribution function of two parameters logistic distribution.

If we choose, $\alpha = 0$, $\beta = 1$, we get standard Logistic distribution. Clearly, the distribution will be —

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

Thus the pdf is, — * $f(x) = \frac{e^{-x}}{(1+e^{-x})^2} = F(x)[1-F(x)]$

We denote, $\rightarrow$ $X \sim L(0,1)$.

■ Expectation:—

$$E(X) = \int_{-\infty}^{\infty} \frac{xe^{-x}}{(1+e^{-x})^2}\, dx$$

$$= \int_{-\infty}^{\infty} \frac{x}{4} \operatorname{sech}^2\left(\frac{x}{2}\right) dx \qquad \left[\text{let, } u = \frac{x}{2},\ du = \frac{1}{2}dx \right]$$

$$= \int_{-\infty}^{\infty} u \operatorname{sech}^2(u)\, du$$

$$= 0 \qquad [\text{odd function}]$$

● Two Parameter Logistic Distribution :—

Definition:— An absolutely continuous random variable X defined over $(-\infty, \infty)$ is said to follow logistic distribution with parameters μ and σ if its p.d.f. is given by,—

$$f(x) = \frac{\frac{1}{\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)}}{\left[1+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]^2}; \qquad -\infty < x < \infty,\ \sigma > 0,\ -\infty < \mu < \infty$$

We write, $X \sim L(\mu, \sigma)$.

In particular, when $\mu = 0$, and $\sigma = 1$, then,—

$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2} I_x(-\infty, \infty).$$

■ Result:— If $X \sim L(\mu, \sigma)$ then $Z = \frac{X-\mu}{\sigma} \sim L(0,1)$.

Proof:— $F_Z(z) = P[Z \le z] = P\left[\frac{X-\mu}{\sigma} \le z\right]$

$$= P[X \le \mu + \sigma z]$$

$$= F_X(\mu + \sigma z)$$

$$\therefore f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X(\mu + \sigma z) = \sigma \cdot f_X(\mu + \sigma z)$$

$$= \frac{e^{-z}}{(1+e^{-z})^2}.$$

■ $$E(X-\mu) = \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)} \frac{dx}{\left[1+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right]^2}$$

$$= \int_{-\infty}^{\infty} \sigma z e^{-z} \frac{dz}{(1+e^{-z})^2} \qquad \left[\text{let, } z = \frac{x-\mu}{\sigma};\ dz = \frac{1}{\sigma}dx\right]$$

$$= \sigma \int_{-\infty}^{\infty} \frac{z e^{z}}{(e^{z}+1)^2}\, dz \qquad \left[\text{let, } g(z) = \frac{ze^z}{(e^z+1)^2},\ g(-z) = -g(z), \text{ i.e. } g(z) \text{ is an odd function}\right]$$

$$= 0$$

$$\Rightarrow E(X) = \mu.$$

MGF of X about mean (μ): —

$$M_{X-\mu}(t) = E\left(e^{t(x-\mu)}\right)$$

$$= \int_{-\infty}^{\infty} e^{t(x-\mu)} \cdot \frac{1}{\sigma} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)} \cdot \frac{dx}{\left\{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}\right\}^2}$$

$$= \int_{-\infty}^{\infty} e^{t\sigma z} \cdot e^{-z} \cdot \frac{dz}{(1+e^{-z})^2}$$

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

$\therefore dx = \sigma dz$

$$= \int_0^1 \left(\frac{u}{1-u}\right)^{t\sigma} du$$

let, $\frac{1}{1+e^{-z}} = u$

$\therefore \frac{e^{-z}}{(1+e^{-z})^2} dz = du$

$$= \beta(t\sigma+1, 1-t\sigma), \text{ provided } |t|<1.$$

$$= \frac{\Gamma(t\sigma+1)\,\Gamma(1-t\sigma)}{\Gamma(2)} = t\sigma\,\Gamma(t\sigma)\,\Gamma(1-t\sigma)$$

$$= t\sigma \cdot \frac{\pi}{\sin \pi t\sigma}.$$

Calculation of moments: —

$$M_{X-\mu}(t) = \frac{t\sigma\pi}{\sin t\sigma\pi} = \frac{\theta}{\sin\theta}, \quad \theta = \pi t\sigma.$$

Now, $\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\cdots\infty$

$$\Rightarrow \sin\theta = \theta\left[1 - \frac{\theta^2}{3!} + \frac{\theta^4}{5!} - \cdots \infty\right]$$

$$\Rightarrow \frac{\theta}{\sin\theta} = [1-\delta]^{-1}, \text{ where } \delta = \frac{\theta^2}{3!} - \frac{\theta^4}{5!} + \frac{\theta^6}{7!} \cdots \infty$$

$$= 1 + \delta + \delta^2 + \delta^3 + \cdots \infty$$

$$= 1 + \left(\frac{\theta^2}{3!} - \frac{\theta^4}{5!} + \frac{\theta^6}{7!} \cdots \infty\right) + \left(\frac{\theta^2}{3!} - \frac{\theta^4}{5!} + \frac{\theta^6}{7!} \cdots\right)^2 + \cdots$$

$$\mu_2 = \text{coefficient of } \frac{t^2}{2!} = \frac{\sigma^2\pi^2}{3}.$$

$$\mu_3 = 0$$

$$\mu_4 = \text{Coefficient of } \frac{t^4}{4!} = \sigma^4\pi^4\left[-\frac{1}{5} + \frac{2}{3}\right]$$

$$= \frac{7}{15}\sigma^4\pi^4$$

$$\therefore \gamma_2 = \frac{\frac{7}{15}\sigma^4\pi^4}{\frac{1}{9}\sigma^4\pi^4} - 3 = \frac{21}{5} - 3 = 1.2 > 0$$

Standard Logistic Distribution: →

If $X \sim \text{Logistic}(\alpha, \beta)$, then $Y = \frac{X-\alpha}{\beta} \sim \text{Logistic}(0,1)$

The DF of Y is $F_Y(y) = \frac{1}{1+e^{-y}}$, $y \in \mathbb{R}$

The PDF of Y is, $f_Y(y) = \frac{e^{-y}}{(1+e^{-y})^2}$, $y \in \mathbb{R}$

Properties: →

i) $f_Y(-y) = \frac{e^{y}}{(1+e^{y})^2} = \frac{e^{-y}}{(1+e^{-y})^2}$

$= f_Y(y)$, $y \in \mathbb{R}$

Hence Y is symmetrically distributed about '0'.

Therefore, the median of Y is at $y=0$.

ii) $F_Y(y) = \frac{1}{1+e^{-y}}$, $y \in \mathbb{R}$

Now, $\frac{F_Y(y)}{1-F_Y(y)} = e^{y}$

$\Rightarrow \ln\left\{\frac{F_Y(y)}{1-F_Y(y)}\right\} = y$

For a logistic distribution with parameters (α, β), i.e.

for $X \sim \text{Logistic}(\alpha, \beta)$,

$\log_e\left\{\frac{F_X(x)}{1-F_X(x)}\right\} = \frac{x-\alpha}{\beta} = -\left(\frac{\alpha}{\beta}\right) + \left(-\frac{1}{\beta}\right)x$.

which is linear in x.

iii) MD of Y about mean $= E|Y| = 2\log_e 2$.

TRUNCATED DISTRIBUTION:

Let X be an absolutely continuous random variable defined over $\mathbb{R}$ with distribution function $F(x)$ and pdf $f(x)$;

Suppose we discard all the values of x for which,

$a < x$ and $x > b$, where $a, b \in \mathbb{R}$

Then the distribution is called a truncated distribution defined over $[a, b]$.

The distribution function of this truncated distribution is given by,

$$F(x/a \le x \le b)$$

$$= P[X \le x / a \le x \le b]$$

$$= \frac{P[(X \le x) \cap (a \le x \le b)]}{P[a \le x \le b]}$$

$$= \begin{cases} 0 & \text{if } x < a \\ \frac{F(x) - F(a)}{F(b) - F(a)} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

Hence, the pdf of truncated distribution is,

$$f(x) = \begin{cases} \frac{f(x)}{F(b) - F(a)} & \text{if } a \le x \le b \\ 0 & \text{ow.} \end{cases}$$

☆ **Problem 1.** If a random variable has the pdf $f(x) = \frac{1}{\theta} \cdot e^{-x/\theta}$, $0 < x < \infty$ then find $E(X/X > \alpha)$ and $V(X/X > \alpha)$, where $\alpha > 0$.

ANS: → If it is given that $x > \alpha$ then the pdf of X will be,

$$f^*(x) = c f(x), \quad \alpha < x < \infty.$$

where, c is such that, — $\int_{\alpha}^{\infty} f^*(x) dx = 1.$

$$\Rightarrow c \left[\int_{\alpha}^{\infty} \frac{1}{\theta} \cdot e^{-x/\theta} dx \right] = 1$$

$$\Rightarrow c \left[-\frac{1}{\theta} \cdot \frac{e^{-x/\theta}}{1/\theta} \right]_{\alpha}^{\infty} = 1$$

$$\Rightarrow c e^{-\alpha/\theta} = 1$$

$$\Rightarrow c = e^{\alpha/\theta}$$

$$\therefore f^*(x) = \frac{1}{\theta} \cdot e^{-(x-\alpha)/\theta} I_x(\alpha, \infty)$$

$$E(X \mid X>\alpha) = \int_{\alpha}^{\infty} x \cdot \frac{1}{\theta} \cdot e^{-(x-\alpha)/\theta}\, dx$$

$$= \int_{0}^{\infty} (\alpha + \theta z) e^{-z}\, dz$$

$$= \alpha + \theta \int_{0}^{\infty} z e^{-z}\, dz$$

$$= \alpha + \theta$$

$$E(X^2 / X>\alpha) = \int_{\alpha}^{\infty} x^2 \cdot \frac{1}{\theta} \cdot e^{-\left(\frac{x-\alpha}{\theta}\right)}\, dx$$

$$= \int_{\alpha}^{\infty} (\alpha + \theta z)^2 \cdot e^{-z}\, dz$$

$$= \int_{0}^{\infty} (\alpha^2 + 2\theta\alpha z + \theta^2 z^2) e^{-z}\, dz$$

$$= \alpha^2 + 2\theta\alpha + \theta^2 \int_{0}^{\infty} z^2 e^{-z}\, dz$$

$$= \alpha^2 + 2\theta\alpha + 2\theta^2$$

$$\therefore V(X / X>\alpha) = \alpha^2 + 2\theta\alpha + 2\theta^2 - (\alpha+\theta)^2$$

$$= \theta^2.$$

☆ Problem 2. Find the mean & varience of truncated normal distribution with pdf

$$f^*(x) = \begin{cases} c \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & \alpha \le x \le \beta \\ 0, & \text{ow.} \end{cases}$$

ANS:-

$$\int_{\alpha}^{\beta} f^*(x) = 1$$

$$\Rightarrow c \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

$$\Rightarrow c \cdot P[\alpha \le Z \le \beta] = 1 \quad [\text{where } Z \sim N(\mu, \sigma^2)]$$

$$\Rightarrow c \cdot P\left[\frac{\alpha-\mu}{\sigma} \le \frac{Z-\mu}{\sigma} \le \frac{\beta-\mu}{\sigma}\right] = 1$$

$$\Rightarrow c \cdot P\left[P\left(\frac{Z-\mu}{\sigma} \le \frac{\beta-\mu}{\sigma}\right) - P\left(\frac{Z-\mu}{\sigma} < \frac{\alpha-\mu}{\sigma}\right)\right] = 1$$

$$\Rightarrow c\left[\Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right)\right] = 1.$$

$$\Rightarrow c = \frac{1}{\Phi(b) - \Phi(a)}, \quad \text{we assume } b = \frac{\beta-\mu}{\sigma}, \; a = \frac{\alpha-\mu}{\sigma}.$$

$$E(x-\mu) = c\int_{\alpha}^{\beta}(x-\mu)\cdot\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$= c\int_{a}^{b}\sigma z\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}(z^2)}dz$$

$$= c\sigma\int_{a}^{b} z\phi(z)dz$$

$$= c\sigma\left[-\phi(z)\right]_a^b$$

$$= c\sigma\left[\phi(a)-\phi(b)\right]$$

$$= \frac{\sigma\left[\phi\left(\frac{\alpha-\mu}{\sigma}\right)-\phi\left(\frac{\beta-\mu}{\sigma}\right)\right]}{\Phi\left(\frac{\beta-\mu}{\sigma}\right)-\Phi\left(\frac{\alpha-\mu}{\sigma}\right)}$$

$$\therefore E(X) = \mu + \sigma c\left[\phi(a)-\phi(b)\right]$$

$$\Rightarrow E^2(X) = \mu^2 + 2\mu\sigma c\left[\phi(a)-\phi(b)\right] + \sigma^2c^2\left[\phi(a)-\phi(b)\right]^2$$

$$\therefore E(X-\mu)^2 = c\int_{\alpha}^{\beta}(x-\mu)^2\cdot\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$= c\int_{a}^{b} z^2\sigma^2\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-z^2/2}dz$$

$$= \sigma^2c\int_{a}^{b} z^2\phi(z)dz$$

$$= \sigma^2c\left[z(-\phi(z))\Big|_a^b + \int_a^b\phi(z)dz\right]$$

$$= \sigma^2c\left[a\phi(a)-b\phi(b)+\left(\Phi(b)-\Phi(a)\right)\right]$$

$$= \sigma^2c\left[a\phi(a)-b\phi(b)+\frac{1}{c}\right]$$

$$= \sigma^2\left[c\left\{a\phi(a)-b\phi(b)\right\}+1\right]$$

$$E(X-\mu)^2 = \sigma^2\left[c\left\{a\phi(a)-b\phi(b)\right\}+1\right]$$

$$E(X^2) - 2\mu E(X)+\mu^2 - E^2(X) = \sigma^2\left[c\left\{a\phi(a)-b\phi(b)\right\}\right]+\sigma^2-E^2(X)$$

$$\Rightarrow V(X) = \sigma^2\left[c\left\{a\phi(a)-b\phi(b)\right\}\right]+\sigma^2-\mu^2+2\mu\left[\mu+\sigma c\left[\phi(a)-\phi(b)\right]\right]$$

$$-\mu^2-2\mu\sigma c\left[\phi(a)-\phi(b)\right]+\sigma^2c^2\left[\phi(a)-\phi(b)\right]^2$$

$$= \sigma^2c\left[\left\{a\phi(a)-b\phi(b)\right\}+c\left\{\phi(a)-\phi(b)\right\}^2\right]+\sigma^2$$

☆ **Problem 3.** Suppose the tails of a normal distribution with mean μ and varience σ^2 are cut off at distance $\pm k\sigma$ $(k>0)$ from μ.

If σ_0^2 be the varience of truncated distribution such that,—

$$\frac{\sigma_0^2}{\sigma^2} = 1 - \frac{ke^{-k^2/2}}{\int_0^k e^{-t^2/2}dt}$$

and also show that a sufficient condition for β_2 of the truncated distribution to be < 3 is that, $k > \sqrt{3}$.

Ans:— After the truncation the pdf of X will be,

$$f(x) = c \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \mu - k\sigma < x < \mu + k\sigma$$

$$c = \frac{1}{\Phi\left(\frac{\mu+k\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-k\sigma-\mu}{\sigma}\right)}$$

$$= \frac{1}{\Phi(k) - \Phi(-k)} = \frac{1}{2\Phi(k)-1}$$

$$= \frac{1}{2\left[\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}dx + \int_0^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx\right] - 1}$$

$$= \frac{1}{\cancel{2}\cdot\frac{1}{\cancel{2}} - 1 + 2\int_0^k \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx}$$

$$= \frac{1}{2\int_0^k \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}dx}$$

$$E(X-\mu) = c\int_{\mu-k\sigma}^{\mu+k\sigma} \frac{(x-\mu)}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

$$= c\int_{-k}^{k} \frac{\sigma z}{\sqrt{2\pi}} \cdot e^{-z^2/2}dz$$

[$\because$ the integrand is an odd function]

$$= c \times 0 = 0$$

$\therefore E(X) = \mu.$

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

$$= c\int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^{2r} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{2c\sigma^{2r}}{\sqrt{2\pi}} \int_0^k z^{2r} e^{-z^2/2} dz$$

$$= 2c\sigma^{2r} \int_0^k z^{2r-1} \cdot z\,\phi(z)\, dz$$

$$= 2c\sigma^{2r}\left[-z^{2r-1} \cdot \phi(z)\Big|_0^k + (2r-1)\int_0^k z^{2r-2} \cdot \phi(z)\, dz\right]$$

$$= 2c\sigma^{2r}\left[-k^{2r-1}\phi(k) + (2r-1)\frac{\mu_{2r-2}}{2c\sigma^{2r-2}}\right]$$

$$= -2c\sigma^{2r}k^{2r-1}\phi(k) + (2r-1)\sigma^2\mu_{2r-2}.$$

Putting, $r=1$,

$$\mu_2 = \sigma_0^2 = -2c\sigma^2 \cdot k \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-k^2/2} + \sigma^2$$

$$\Rightarrow \frac{\sigma_0^2}{\sigma^2} = 1 - 2ck \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-k^2/2}$$

$$= 1 - \frac{ke^{-k^2/2}}{\int_0^k e^{-x^2/2} dx} \quad \text{[Putting value of } c\text{]}$$

Putting, $r=2$,

$$\mu_4 = 3 - \frac{6ke^{-k^2/2}}{2\Phi(k)-1} - \frac{6k^3e^{-k^2/2}}{2\Phi(k)-1}$$

Hence, $$\beta_2 = \left\{3 - \frac{6ke^{-k^2/2}}{2\Phi(k-1)}(1+k^2)\right\} \Bigg/ \left\{1 - \frac{2ke^{-k^2/2}}{\int_0^k e^{-u^2/2} du}\right\}^2$$

$\therefore \beta_2 < 3$ in case $k^2 > 3$ i.e. $k > \sqrt{3}$.

TRUNCATED DISTRIBUTION

Let X be a RV and let $\mathscr{X}$ be the sample space such that $P[X \in \mathscr{X}] = 1$.

Let $A = (a, b] \subset \mathscr{X}$.

We wish to find the probability distribution of X over $A = (a, b]$. Let $F(x)$ be the CDF of X over $\mathscr{X}$. Then CDF of X over $A = (a, b]$ is given by,

$$G(x) = P[X \le x \mid x \in A]$$

$$= P[X \in A]$$

$$= \begin{cases} \dfrac{P[a < X \le x]}{P[a < X \le b]} & , \text{if } x \in A \\ 0 & , \text{if } x \le a \\ 1 & , \text{if } x > b \end{cases}$$

$$= \begin{cases} \dfrac{F(x) - F(a)}{F(b) - F(a)} & , \text{if } x \in A = (a, b] \\ 0 & , \text{if } x \le a \\ 1 & , \text{if } x > b \end{cases}$$

The probability distribution of X over $A \subset \mathscr{X}$, is called a truncated distribution of X.

i) <u>Discrete Case:</u> — Let X be a discrete RV with PMF $p(x)$ then the PMF of X over A is

$$g(x) = G(x) - G(x-0) \quad , \text{if } x \in A$$

$$= \frac{F(x) - F(x-0)}{P[X \in A]}$$

$$= \begin{cases} \dfrac{p(x)}{P[X \in A]} & , \text{if } x \in A \\ 0 & , \text{OW} \end{cases}$$

ii) <u>Continuous Case:</u> — Let X be a continuous R.V. with PDF $f(x)$ then the PDF of X over A is

$$g(x) = G'(x) = \begin{cases} \dfrac{f(x)}{P[X \in A]} & , x \in A \\ 0 & , \text{OW} \end{cases}$$

Also note that, $\displaystyle\int_a^b f(x)\,dx = P[X \in A]$

$$\Rightarrow \int_a^b \frac{f(x)\,dx}{P[X \in A]} = 1$$

This implies $g(x) = \dfrac{f(x)}{P[X \in A]}$, $x \in A$, defines a PDF, indeed.

Some selected Truncated distribution :

[A]. Let $X \sim Bin(n, p)$. Find the PMF of X over $A = \{1, 2, 3, \ldots, n\}$. Also, find the mean and variance of the truncated distribution. Find the PGF also.

Soln. → The PMF of X is
$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & , x = 0(1)n,\ q = 1-p. \\ 0 & , \text{OW} \end{cases}$$

Here, $\mathscr{X} = \{0, 1, 2, \ldots, n\}$ and $A = \{1, 2, 3, \ldots, n\}$

The PMF of X over $A = \{1, 2, \ldots, n\}$ is
$$g(x) = \begin{cases} \dfrac{p(x)}{P[X \in A]} & , \text{if } x \in A \\ 0 & , \text{OW} \end{cases}$$

$$= \begin{cases} \dfrac{\binom{n}{x} p^x q^{n-x}}{1-q^n} & , x = 1(1)n \\ 0 & , \text{OW} \end{cases}$$

$[\because P[X \in A] = 1 - P[X = 0] = 1 - q^n.]$

The mean of the truncated distribution is
$$E(X/X \in A) = \sum_{x=1}^{n} x . g(x) = \sum_{x=1}^{n} \frac{x \binom{n}{x} p^x q^{n-x}}{(1-q^n)}$$
$$= \frac{E(X)}{1-q^n} = \frac{np}{1-q^n}.$$

Now, $V(X | X \in A) = E\{X(X-1)/X \in A\} + E\{X/X \in A\} - E^2(X/X \in A)$
$$= \frac{\sum_{x=1}^{n} x(x-1) \binom{n}{x} p^x q^{n-x}}{1-q^n} + \left(\frac{np}{1-q^n}\right) - \left(\frac{np}{1-q^n}\right)^2$$
$$= \frac{n(n-1)p^2 + np}{1-q^n} - \left(\frac{np}{1-q^n}\right)^2.$$

PGF is,
$$P(t) = \sum_{x=1}^{n} t^x g(x) = \frac{\sum_{x=0}^{n} t^x \binom{n}{x} p^x q^{n-x} - q^n}{(1-q^n)}$$
$$= \frac{(q+pt)^n - q^n}{(1-q^n)}.$$

$\therefore$ The PGF of the truncated distribution is $= \dfrac{(q+pt)^n - q^n}{(1-q^n)}$.

Scanned by CamScanner

B. Let X be a $P(\lambda)$ variable but only the values $1, 2, 3, \ldots\ldots$ are recorded, i.e. all values are recorded except $x = 0$. Find the distn. of X over $A = \{1, 2, 3, \ldots\}$. Also find the mean and variance of the distribution. Find the PGF of X also.

Soln. → The PMF of X is

$$p(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x!} & , x = 0, 1, 2, \ldots \\ 0 & , \text{ow} \end{cases}$$

Here, $\mathcal{X} = \{0, 1, 2, \ldots\}$ and $A = \{1, 2, 3, \ldots\}$

The PMF of the truncated distn. is,

$$g(x) = \begin{cases} \dfrac{e^{-\lambda} \cdot \lambda^x}{x! \, (1 - e^{-\lambda})} & , x = 1, 2, 3, \ldots \\ 0 & , \text{ow}. \end{cases}$$

$$E(X / X \in A) = \frac{\sum_{x=1}^{\infty} x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!}}{1 - e^{-\lambda}}$$

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

$$\text{var}(X / X \in A) = E[X(X-1) / X \in A] + E[X / X \in A] - E^2(X / X \in A)$$

$$= \frac{\sum_{x=1}^{\infty} x(x-1) \frac{e^{-\lambda} \cdot \lambda^x}{x!}}{1 - e^{-\lambda}} + \left(\frac{\lambda}{1 - e^{-\lambda}}\right) - \left(\frac{\lambda}{1 - e^{-\lambda}}\right)^2$$

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

PGF of X over A is,

$$P(t) = \sum_{x=1}^{\infty} t^x g(x) = \sum_{x=1}^{\infty} t^x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} \Big/ (1 - e^{-\lambda})$$

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

$$= \frac{e^{-\lambda}(e^{\lambda t} - 1)}{1 - e^{-\lambda}}.$$

Now,

$$P'(t) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \{\lambda e^{\lambda t} - 0\}$$

$$P''(t) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \{\lambda^2 \cdot e^{\lambda t}\}$$

$$\therefore E\{X / X \in A\} = P'(t)\big|_{t=1} = \frac{\lambda}{1 - e^{-\lambda}},$$

$$\text{and } E\{X(X-1) / X \in A\} = P''(t)\big|_{t=1} = \frac{\lambda^2}{1 - e^{-\lambda}}.$$

[C]. Let $X \sim$ Cauchy $(0,1)$. Find the truncated distn. of X over $A = (-\beta, \beta)$. Hence find the mean, if exists, of the distn. Describe a use of the truncated distribution.

Soln $\rightarrow$ The PDF is, $f(x) = \dfrac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$

Here $\mathcal{X} = \mathbb{R}$ and $A = (-\beta, \beta)$.

The PDF of X over $A = (-\beta, \beta)$ is, —

$$g(x) = \begin{cases} \dfrac{f(x)}{P[X \in A]} & , x \in A \\ 0 & , \text{ow} \end{cases}$$

$$= \begin{cases} \dfrac{\frac{1}{\cancel{\pi}(1+x^2)}}{\frac{2}{\cancel{\pi}}\tan^{-1}\beta} & , x \in (-\beta, \beta) \\ 0 & , \text{ow} \end{cases}$$

$$\left[P(X \in A) = \int_{-\beta}^{\beta} \frac{dx}{\pi(1+x^2)} = \frac{1}{\pi}\left[\tan^{-1} x\right]_{-\beta}^{\beta} = \frac{2}{\pi}\tan^{-1}\beta \right]$$

$$= \begin{cases} \dfrac{1}{2(1+x^2)\tan^{-1}\beta} & , -\beta < x < \beta \\ 0 & , \text{ow} \end{cases}$$

Note that, $$\int_{-\beta}^{\beta} |x|\, g(x)\, dx = \frac{1}{2\tan^{-1}\beta} \int_{-\beta}^{\beta} \frac{|x|}{1+x^2}\, dx$$

$$= \frac{1}{2\tan^{-1}\beta} \int_{0}^{\beta} \frac{2x}{1+x^2}\, dx$$

$$= \frac{1}{2\tan^{-1}\beta} \cdot \log(1+\beta^2) < \infty$$

Hence, the mean of the truncated distribution exists.

Now, $$E(X / X \in A) = \int_{-\beta}^{\beta} x\, g(x)\, dx$$

$$= \frac{1}{2\tan^{-1}\beta} \int_{-\beta}^{\beta} \frac{x}{1+x^2}\, dx = 0$$

The mean of the cauchy distribution does not exist due to the presense of thick tails.

The problem of non-existance of mean can be removed by the truncation of the tails.

D. Let X follows Exponential distribution with mean $\frac{1}{\theta}$.
Find the distribution of X truncated, —
(i) above 'a', (ii) below 'a',
Also find the mean and variance of the truncated distribution.

soln. → The PDF of X is $f(x) = \begin{cases} \theta e^{-\theta x}, & \text{if } x>0 \\ 0, & \text{OW} \end{cases}$

i) Here $\mathcal{X} = (0,\infty)$ and $A = [a,\infty)$
The PDF of X over $A = [a,\infty)$ is

$$g(x) = \begin{cases} \dfrac{\theta e^{-\theta x}}{e^{-\theta a}}, & \text{if } x \geq a \\ 0, & \text{OW} \end{cases} \quad \left[P[X\in A] = \int_a^\infty \theta e^{-x\theta} dx = \frac{\theta.[-e^{-x\theta}]_a^\infty}{\theta} = e^{-a\theta}. \right]$$

$$= \begin{cases} \theta . e^{-\theta(x-a)}, & \text{if } x \geq a \\ 0, & \text{OW} \end{cases}$$

which is the PDF of shifted exponential distribution.
The mean of the truncated distribution is

$$= E[X/X\geq a] \text{ and, } E[X-a/X\geq a] = \int_a^\infty (x-a)\, g(x) dx$$

$$= \int_a^\infty (x-a).\theta.e^{-\theta(x-a)} dx$$

$$= \frac{1}{\theta}\int_0^\infty z.e^{-z} dz, \text{ where } z = \theta(x-a)$$

$$= \frac{\Gamma(2)}{\theta}$$

$$= \frac{1}{\theta}.$$

$$\therefore E(X/X\geq a) = a + \frac{1}{\theta} = \text{mean}.$$

$$\therefore E(X^2/X\geq a) = \int_a^\infty x^2.\frac{1}{\theta}.e^{-\left(\frac{x-a}{\theta}\right)} dx$$

$$= \int_a^\infty (a+\theta z)^2 e^{-z} dz \qquad \left[\text{let, } z = \frac{x-a}{\theta}\right]$$

$$= a^2 + 2\theta a + \theta^2 \int_0^\infty z^2 e^{-z} dz$$

$$= a^2 + 2\theta a + 2\theta^2$$

$$\therefore V(X/X\geq a) = a^2 + 2\theta a + 2\theta^2 - (a+\theta)^2$$

$$= \theta^2.$$

$\boxed{E}$. Let $X \sim N(\mu, \sigma^2)$. Find the PDF of X over $A = (a, b)$. Also find the mean and variance of the distribution.

Soln. $\rightarrow$ The PDF of X is; $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $x \in \mathbb{R}$

$$= \frac{1}{\sigma} \Phi\left(\frac{x-\mu}{\sigma}\right), \; x \in \mathbb{R}$$

Here $\mathcal{X} = \mathbb{R}$ and $A = (a, b)$

The PDF of X over $A = (a, b)$ is

$$g(x) = \begin{cases} \frac{f(x)}{P[X \in A]} & , \text{if } x \in A \\ 0 & , \text{ow} \end{cases}$$

where, $P[X \in A] = P[a < X < b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$

Note that,

$$E\left(\frac{X-\mu}{\sigma} / X \in A\right) = \int_A \left(\frac{x-\mu}{\sigma}\right) g(x)\,dx$$

$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} z \cdot \frac{\phi(z)}{P[X \in A]}\,dz, \text{ where } z = \frac{x-\mu}{\sigma}.$$

$$= \frac{1}{P[X \in A]} \left[-\phi(z)\right]_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \qquad \left[\because \int z\phi(z)\,dz = -\phi(z) + c\right]$$

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

Hence, $E\left[X / X \in A\right] = \mu + \sigma \cdot E\left\{\left(\frac{X-\mu}{\sigma}\right) / X \in A\right\}$

$= \mu + \sigma\theta$, where, $\theta = \frac{\phi(a^*) - \phi(b^*)}{\Phi(b^*) - \Phi(a^*)}$

$a^* = \frac{a-\mu}{\sigma}$ and $b^* = \frac{b-\mu}{\sigma}$.

$$E\left\{\left(\frac{X-\mu}{\sigma}\right)^2 / X \in A\right\} = \int_A \left(\frac{x-\mu}{\sigma}\right)^2 g(x)\,dx = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{z^2\phi(z)}{P[X \in A]}\,dz$$

$$= \frac{1}{P[X \in A]} \left\{\left[-z\phi(z)\right]_{a^*}^{b^*} + \int_{a^*}^{b^*} \phi(z)\,dz\right\}$$

$$= \frac{a^*\phi(a^*) - b^*\phi(b^*)}{\Phi(b^*) - \Phi(a^*)} + 1$$

Scanned by CamScanner

$$\therefore \operatorname{var}(X/X\in A) = \operatorname{var}\{(X-\mu)/X\in A\}$$

$$= \sigma^2 \left[E\left\{ \left(\frac{X-\mu}{\sigma}\right)^2 / X\in A \right\} - E^2\left\{ \frac{X-\mu}{\sigma} / X\in A \right\} \right]$$

$$= \sigma^2 \left\{ \frac{a^*\phi(a^*) - b^*\phi(b^*)}{\Phi(b^*) - \Phi(a^*)} + 1 - \theta^2 \right\}$$

[In particular,

if $A = [a, \infty)$, then,

$$E[X/X\in A] = \mu + \sigma\theta, \text{ where } \theta = \frac{\phi(a^*)}{1-\Phi(a^*)},$$

$$a^* = \frac{a-\mu}{\sigma}.$$

and $\operatorname{var}(X/X\in A) = \sigma^2\{a^*\theta + 1 - \theta^2\}$]

× ——— ×

Scanned by CamScanner

☆ **Problem:-** Let $X \sim N(\mu, \sigma^2)$ and $A = (\mu - k\sigma, \mu + k\sigma)$, $k > 0$. Find the PDF of X over A. Also, find the mean and variance of truncated distribution.

Soln. $\Rightarrow$ The PDF of X is, $f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$, $x \in \mathbb{R}$

Now,
$$P[X \in A] = P[\mu - k\sigma < X < \mu + k\sigma]$$
$$= P\left[-k < \frac{X-\mu}{\sigma} < k\right]$$
$$= \int_{-k}^{k} \phi(t)\,dt, \quad \text{where } t = \frac{x-\mu}{\sigma},\ dt = \frac{1}{\sigma}dx.$$
$$= \frac{2}{\sqrt{2\pi}} \int_0^k e^{-t^2/2}\,dt$$
$$= \sqrt{\frac{2}{\pi}} \cdot I(k), \text{ say},$$

The PDF of X over A is,
$$g(x) = \begin{cases} \dfrac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\sqrt{\frac{2}{\pi}} \cdot I(k)}, & x \in (\mu - k\sigma, \mu + k\sigma) \\ 0, & \text{ow} \end{cases}$$

Note that, $g(\mu - h) = g(\mu + h)$ $\forall\, h$

$\Rightarrow$ The truncated distribution is symmetric about μ.

Hence,
$$E(X / X \in A) = \mu, \text{ and}$$
$$\mu_{2n-1}(A) = E\left\{(X-\mu)^{2n-1} / X \in A\right\}$$
$$= 0, \quad n \in \mathbb{N}$$

Now,
$$\text{Var}[X / X \in A] = E\left\{(X-\mu)^2 / X \in A\right\} = v^2$$
$$= \int_{-k}^{k} (x-\mu)^2 g(x)\,dx$$
$$= \frac{\sigma^2}{\sqrt{\frac{2}{\pi}} \cdot I(k)} \int_{-k}^{k} z^2 \cdot \phi(z)\,dz, \quad z = \frac{x-\mu}{\sigma}.$$
$$= \frac{2\sigma^2}{\sqrt{\frac{2}{\pi}} \cdot I(k)} \int_0^k z^2 \cdot \phi(z)\,dz$$

$$= \frac{2\sigma^2}{\sqrt{\frac{2}{\pi}}\cdot I(k)} \left\{ \left[-z\phi(z)\right]_0^k + \int_0^k \phi(z)\,dz \right\}$$

$$= \frac{\frac{2}{\sqrt{2\pi}}\cdot\sigma^2}{\sqrt{\frac{2}{\pi}}\cdot I(k)} \left\{ -ke^{-k^2/2} + I(k) \right\}$$

$$= \frac{\sigma^2}{I(k)} \left\{ I(k) - ke^{-k^2/2} \right\}$$

$$\therefore \quad \left\{ 1 - \frac{v^2}{\sigma^2} \right\} I(k) = k\cdot e^{-k^2/2}$$

Here, $\gamma_1(X/A) = 0$, as $\mu_3(A) = 0$.

Now, $\mu_4(A) = E\left[(X-\mu)^4 / X \in A \right]$

$$= \frac{2\sigma^4}{\sqrt{\frac{2}{\pi}}\cdot I(k)} \int_0^k z^4 \phi(z)\,dz$$

$$= \frac{\sigma^4}{I(k)} \left[3I(k) - k^3\cdot e^{-k^2/2} - 3k\cdot e^{-k^2/2} \right]$$

and $\gamma_2(X/A) = \dfrac{\mu_4(A)}{\mu_2^2(A)} - 3$

$$= \frac{-k\cdot e^{-k^2/2}\left\{ (k^2-3)\cdot I(k) + 3\cdot k\cdot e^{-k^2/2} \right\}}{\left\{ I(k) - k\cdot e^{-k^2/2} \right\}^2}$$

If $k > \sqrt{3}$, then $\gamma_2(X / X \in A) < 0$

and the truncated distribution is platykurtic.

SCALING METHODS

Unlike physical or biological characteristics ~~phy~~ phsycological characteristics are rather abstract and hence can be measured only with same degree of unreliability. In phsycology and education characters under study are ~~phy~~ phsycological traits or mental abilities like intelligence, aptitude, interest, opinion, atitude or simply scholastic achievement. these characters can be measured by performing experiments over entities by subjecting them to several tests. Different tests involve different degree of mental ability and performance. Here the most practical consideration is that the scales of different tastes should be comparable. For the purpose of measurement of these traits one has to develop a certain scale which bears a strong analogy with an ordinary scale for measuring lengths. Equal distances on a phsycological scale should stand for empirically equal differences in the phsycological trait being measured.

But the '0' point on the phsycological scale is arbitary. However, distances from the arbitary 0 point should be additive. Thus, a phsycological scale is an interval scale and not a ratio scale. In this scholastic test often a total score is required for an individual who is appearing in several test. the present procedure of additing the raw scores on all the tests to get a total score for the individual is not a a valid one. Since the same raw score x on different tests may involve different degrees of ability and hence may not be equivalent in different tests. Hence, the raw scores has to be scaled under some assumptions regarding the distribution of the trait which the test is measuring.

[P.T.O]

Scanned by CamScanner

i) Percentile Scaling :— Here we assume that the distribution of the trait under consideration is rectangular, under which we shall have percentile differences equal throughout the scale. To determine the scale value corresponding to a score X on a test we have to find the percentile position of an individual with score x, i.e. percentage of individuals in the group having a score equal to or less than x. For e.g. a person having a percentile rank of 30 is situate above 30% of the group of which he is a member. This can be easily obtained from the score distribution, assuming that the score is a continuous variable. Irrespective of the form of the original raw scores distribution, percentile scores reflect the comparative performance of an individual w.r.t. all the individuals in the group and doesn't reflect the individual's absolute ability. The distribution of raw scores is rarely rectangular so that the basic underlying, the percentile scaling may not be always realistic. Thus by using the scaling method one should be aware of its limitations.

ii) Z - scaling :— Here we assume that the raw scores of all the tests have the same distribution, but they differ only in mean and standard deviation. Thus if the scores X and Y on two subjects are distributed with means μ_1 and μ_2 and s.d.'s σ_1 and σ_2, respectively then the standardised scores are $\frac{X-\mu_1}{\sigma_1}$ and $\frac{Y-\mu_2}{\sigma_2}$ and they can be compared. In particular, if the mean is arbitrarily taken to be zero and the s.d. to be unity, the scores are called standard scores or σ scores or z-scores. To avoid negative standard scores, the mean is generally taken as 50 and s.d. as 10 in linear derived scores. If a particular test has a raw score mean μ and s.d. σ, then the linear derived score w corresponding to a raw score x on that test is given by $\frac{X-\mu}{\sigma} = \frac{w-50}{10}$

$$\Rightarrow w = \underbrace{50 + 10\left(\frac{X-\mu}{\sigma}\right)}_{\text{linear derived scores}}$$

$\therefore w = 50 + 10z$, where z is the standard score and w is the linear derived score with mean 50 and s.d. 10. This linear transformation changes only the mean and s.d., while retaining the form of the original distribution.

iii) **T-score :—** In this case, we assume that the trait distribution is normal with mean 50 and s.d. 10.

To get the T-score corresponding to a raw score x, we first find the percentile position p of an individual with score x and find the value T of a normal distn. with mean 50 and s.d. 10, below which the area is $\frac{p}{100}$.

$$T \sim N(50, 100)$$

$$\Phi\left(\frac{T-50}{100}\right) = \frac{p}{100}$$

$$P[T \le t] = \frac{p}{100}$$

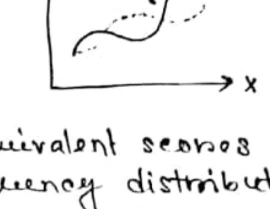
$$P\left[\frac{T-50}{100} \le \frac{t-50}{100}\right] = \frac{p}{100}$$

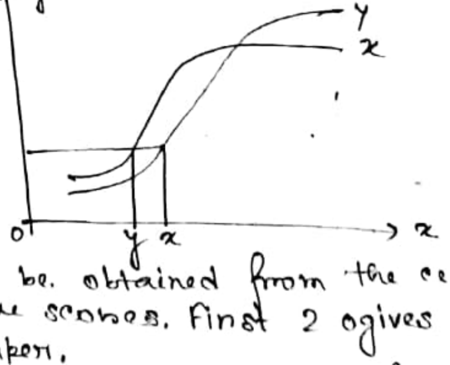
$$\Phi\left(\frac{t-50}{10}\right) = \frac{p}{100}$$

Normalised scores and percentile scores are merely special cases of non-linear transformation of raw scores. A transformation is non-linear if it changes the form of the distribution. For non-linear transformation any form of the distribution may be chosen.

iv) **Method of Equivalent scores :—** Here we don't make any assumption about the trait under consideration. Here we transform the score of one test in terms of that of the other and compare them. Let X and Y be the raw scores on 2 tests with PDFs f_X and g_Y, respectively (obtained by some process of graduation). Two scores x_i of X and y_i of Y are considered equivalent iff $\int_{-\infty}^{x_i} f_X(x)\,dx = \int_{-\infty}^{y_i} g_Y(y)\,dy$

In practise, an equivalence curve may be obtained by computing a number of pairs of equivalent scores (x_i, y_i) and fitting to the corresponding set of points an appropriate curve $y = h(x)$.





Equivalent scores can also be obtained from the cumulative frequency distribution of the scores. First 2 ogives are drawn on the same graph paper.

In this method, the form of the distribution of equivalent scores is same as that of the standard test. If however the standard test score has a normal distribution the method reduces to normalised scaling.

Scanned by CamScanner

v) **Likert's Scaling** :— Suppose there are 2 judges, rating a group of individuals for their phsycological traits like honesty, responsibility, tactfulness, etc. the frequency distributions of ratings of the 2 judges is known. The problem is to assign weights or numerical scores to the ratings so that the ratings of the 2 judges may be compared or combined. Let us assume that the trait under consideration is normally distributed with mean zero and standard distribution unity. Now, suppose that the individuals with trait values from X_1 to X_2 are given a particular rating. The scale value of this rating is taken to be the mean trait value of all these individuals and is given by.

$$\frac{\int_{x_1}^{x_2} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx}{\int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx} = \frac{\phi(x_1) - \phi(x_2)}{\Phi(x_2) - \Phi(x_1)}$$

This is known as Likert's scaling and this method is known as Category scale method.

Thurstone's product scale :— It often happens that the trait in which we are interested can't be expressed as test scores. In such cases excellence of performance is determined by comparing an individual's product with various standard products, the values of which are determined by a number of judges. Suppose there are K standard products judged by a group of N judges. Each of the $\binom{K}{2}$ pairs of products is presented to each judge who selects one member of each pair in reference to the other. the data can be presented in the form of proportion matrix

$$\begin{array}{cc|cccc} & & & \text{Product} & & \\ & & 1 & 2 & \cdots & K \\ \hline & 1 & p_{11} & p_{21} & \cdots & p_{K1} \\ & 2 & p_{12} & p_{22} & \cdots & p_{K2} \\ \text{Product} & \vdots & \vdots & \vdots & & \vdots \\ & K & p_{1K} & p_{2K} & \cdots & p_{KK} \end{array}$$

Here p_{ij} is the proportion of cases in which product i to product j. $p_{ij} = 1 - p_{ji}$

$p_{ii} = \frac{1}{2}$.

Suppose that the distribution of the difference in judgements T_{ij} of the i^{th} and j^{th} product is normal $T_{ij} \sim N(s_i - s_j, \sigma_{i-j}^2)$, where $s_i - s_j$ is the difference of their scale values.

$$P_{ij} = P(T_{ij} > 0) = \frac{1}{\sigma_{i-j}\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}\left[\frac{T_{ij} - (s_i - s_j)}{\sigma_{i-j}}\right]^2}$$

$$= 1 - \Phi\left(-\frac{s_i - s_j}{\sigma_{i-j}}\right).$$

Assuming that the distribution of judgement for each product has the same s.d. σ and that the judgements for any products are uncorrelated.

$$\sigma_{i-j} = \sigma\sqrt{2}$$

Taking $\sigma_{i-j} = \sigma\sqrt{2}$ as the unit of scale and P_{ij} as estimated by p_{ij}, we have.

$$1 - p_{ij} = \Phi(\hat{s}_j - \hat{s}_i)$$

$$\hat{s}_i - \hat{s}_j = -\Phi^{-1}(1 - p_{ij})$$

$$= -x_{ij}$$

Thus, we get $s_i - s_j$ matrix

	1	2	k
1	$s_1 - s_1$	$s_2 - s_1$	$s_k - s_1$
2	$s_1 - s_2$	$s_2 - s_2$	$s_k - s_2$
⋮	⋮	⋮	⋮
k	$s_1 - s_k$	$s_2 - s_k$ ...	$s_k - s_k$
	$s_1 - \bar{s}$	$s_2 - \bar{s}$	$s_k - \bar{s}$

If we take the origin at $\bar{s}$, then the column means provide us a scale value for the k products.

<u>Note:</u> $\frac{ks_1 - \sum s_i}{k} = s_1 - \bar{s}$

If $\bar{s} = 0$, then we get s_1.

Scanned by CamScanner

SOME SCALING PROCEDURE:

A new discipline called 'Psychometry' has been developed as a branch of psychology which deals with the measurement of psychological ~~traits~~ on the mental abilities like intelligence, aptitude, opinion, etc.

For the measurement of psychological and educational characteristics which are rather abstruct in nature as compared with physical or biological characteristics. For the purpose of measurement, one has to develop a certain scale, and the most practical consideration is that the scales for different tests should be comparable. The zero point of psychological scale is arbitrary, the distances from arbitrary zero are additive. In other words, psychological scale is an interval scale and not a ratio scale, since there is no absolute zero point.

Scaling Procedures:

We shall discuss some of the common scaling procedures used in psychology and education:

(i) Scaling individual test items in terms of difficulty.

(ii) Scaling of scores on a test: T-scaling or normalised score; Method of equivalent of scores.

(iii) Scaling of ranking in terms of normal probability curve.

In the scaling procedures developed here, it is assumed that the trait under consideration is normally distributed.

(i) Scaling individual test-items in terms of difficulty:

In this case a number of test items (or problems), all designed to a test, are administered to a large group of individuals who are selected at random out of those for whom the final test is intended and we are interested in arranging these items (or problems) in order of difficulty. For this, the set of problems is given to a large group of individuals for solving them and for each problem (or item) the proportion of those who could solve it is obtained. Let p_i be the proportion of individuals solving the i^{th} problem, $i = 1(1)n$. Of course, the larger the percentage of people (i.e. $100p_i$) passing a test item i, the lower it is of order of difficulty. In the construction of the difficulty scale we assume that the ability on the trait (X) being measured follows $N(\mu, \sigma^2)$ distn.

WLG, we can assume $\mu=0$. The difficulty value of an item is usually defined as the minimum ability to answer this item correctly under the assumption that the ability is distributed Normally $N(0, \sigma^2)$. Therefore, we have $P[X > \sigma x_i] = p_i$, hence $x_i\sigma$ is the amount of ability required for passing the item & may be taken as a measure of difficulty (d_i) for the i^{th} item.

For given p_i's, the values of x_i's can be read from the table of areas under standard normal probability curve.

(ii) Scaling test-scores in several tests :—

Suppose a number of candidates are given five different tests, say, English, Bengali, Physics, Chemistry and Statistics. The question arises: Are we justified in making comparisons on the basis of the sum of raw scores? The answer is 'NO'; since the same raw scores x (say) in different tests, e.g., English and Statistics may require different degrees of ability and hence may not be equivalent and therefore can't be compared meaningfully. In order to make valid comparisons between the raw scores, we need a common scale which is obtained under some assumption regarding the trait distribution.

(a) Normalised Scores / T-Scaling :→ Here we assume that the trait (X) distn. is $N(\mu_X, \sigma_X^2)$ and the raw scores are converted into a system of Normalised scores by transforming into the equivalent points of a normal distn.

Let p be the proportion of individuals getting scores $\le x$. Then

$$p = P[X \le x] = P\left[Z \le \frac{x-\mu_X}{\sigma_X} = \xi, \text{say}\right], \text{ where } Z \sim N(0,1)$$
$$= \Phi(\xi).$$

The number ξ obtained from $\Phi(\xi) = p$, is called the <u>normalised score</u> corresponding to a raw score x.

For practical convenience, normalised scores are transformed to new scale with mean μ and s.d. (σ), by the relation:

$$\frac{\eta-\mu}{\sigma} = \xi \quad \text{———} (*)$$

$\Rightarrow \eta = \mu + \sigma\xi$, where μ, σ are preassigned values, η's are called <u>normalised standard scores.</u>

In particular, if we take $\mu = 50$, $\sigma = 10$ in $(*)$, we get <u>T-scores</u>. Thus T-scores are normalised scores, converted into a distribution with mean 50 and s.d. 10 and are given by : $T = 50 + 10\xi$.

(b) Method of equivalent scores :– Here we do not make any assumption about the trait distn. . The appropriate trait distn is obtained by graduating the raw score distribution. Let X and Y be the scores on two tests having p.d.f.s $f(x)$ and $h(y)$, respectively, obtained by some process.

$$\int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{y} h(t)\,dt \Leftrightarrow F_X(x) = F_Y(y), \text{ where,}$$

$F_X(\cdot)$ and $F_Y(\cdot)$ the c.d.f's of X and Y. Then the curve, say, $y=g(x)$, obtained by solving $F_X(x) = H_Y(y)$, is called equivalence curve of Y for given values of X.

For practical convenience, an equivalence curve may be obtained as follows : First, two ogives are drawn on the same graph paper. Two scores x and y with the same relative cumulative frequency are then regarded as equivalent. Compute a number of pairs of equivalent scores (x_i, y_i), $i=1(1)n$ and fitting to an appropriate curve to the points (x_i, y_i), we get an equivalence curve $y=g(x)$.

For the purpose of comparison or combination, the raw scores on different tests may be converted onto equivalent scores on a standard test.

EXAMPLE :–

1. Find an equivalence curve of two traits X and Y
a) if both X and Y have Rectangular with parameters θ_1 and θ_2.
b) if both X and Y have Exponential with parameters θ_1 and θ_2
Also comment on the results.

Solution :– By definition, two scores x and y on two tests are equivalent iff $\int_{-\infty}^{x} f_X(t)\,dt = \int_{-\infty}^{y} f_Y(t)\,dt$, where $f_X(\cdot)$ and $f_Y(\cdot)$ are the p.d.f.'s of X and Y.

(a) $X \sim Rec(0, \theta_1)$, $Y \sim Rec(0, \theta_2)$ $\Rightarrow$ $\begin{cases} f_X(t) = \begin{cases} \frac{1}{\theta_1}, & 0<t<\theta_1 \\ 0, & \text{ow} \end{cases} \\ f_Y(t) = \begin{cases} \frac{1}{\theta_2}, & 0<t<\theta_2 \\ 0, & \text{ow} \end{cases} \end{cases}$

Hence,
$$\int_0^x \frac{1}{\theta_1}\,dt = \int_0^y \frac{1}{\theta_2}\,dt$$
$$\Rightarrow \frac{x}{\theta_1} = \frac{y}{\theta_2}$$
$$\Rightarrow y = \frac{\theta_2}{\theta_1}x.$$

Hence, the line $y = \frac{\theta_2}{\theta_1}x$ passing through the origin, is the equivalence curve.

(b) $X \sim Exp(\theta_1)$, $Y \sim Exp(\theta_2)$.

Hence, $\int_0^x \theta_1^{-1} e^{-\theta_1^{-1}t}\,dt = \int_0^y \theta_2^{-1} e^{-\theta_2^{-1}t}\,dt$

$$\Rightarrow \left[-e^{-\theta_1^{-1}t}\right]_0^x = \left[-e^{-\theta_2^{-1}t}\right]_0^y$$

$$\Rightarrow \theta_2^{-1}y = \theta_1^{-1}x \Rightarrow y = \frac{\theta_2}{\theta_1}x$$

Hence, in both cases, we have the same equivalence curve $y = \frac{\theta_2}{\theta_1}x$.

2. If the traits $X \sim N(0,1)$ and $Y \sim Cauchy(0,1)$ and if x, y are equivalent scores on the two tests, then show that

$$y = \tan\left[\left\{\Phi(x) - \frac{1}{2}\right\}\pi\right].$$

Solution:- By definition of equivalent scores on two tests, we have

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt = \int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\,dt$$

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

$$\Rightarrow \Phi(x) = \frac{1}{\pi}\tan^{-1}(y) - \frac{1}{\pi}\left(-\frac{\pi}{2}\right)$$

$$\Rightarrow y = \tan\left[\left\{\Phi(x) - \frac{1}{2}\right\}\pi\right]$$

3. If the traits $X \sim N(0,1)$ and $Y \sim$ Double Exponential, for two tests, then find the equivalence curve of y and x.

Solution:- Let x and y be the two equivalent scores on two tests, then

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt = \int_{-\infty}^{y} \frac{1}{2} e^{-|t|}\,dt$$

$$\Rightarrow \Phi(x) = \int_{-\infty}^{0} \frac{1}{2} e^{t}\,dt + \int_{0}^{y} \frac{1}{2} e^{-t}\,dt$$

$$= \frac{1}{2} + \frac{1}{2}\left[-e^{-t}\right]_0^y$$

$$= 1 - \frac{1}{2}e^{-y}.$$

$$\Rightarrow e^{-y} = 2\{1 - \Phi(x)\}$$

$$\Rightarrow y = -\ln\left[2\{1 - \Phi(x)\}\right]$$

(iii) Scaling of rating or ranking in terms of the normal curve :—

(a) Scaling of rating :— Let us suppose that a group of individuals have been rated by different judges w.r.t. some trait, say, honesty and the corresponding frequency distribution of the ratings of the judges given or known. The problem is : "Can we assign weights or numerical scores to these ratings so we can make them comparable from judge to judge ?" The answer is 'yes', provided that we are justified in assuming (i) the normality of the trait distn..

Let us suppose that the distn. of the trait $(X) \sim N(0,1)$. Suppose that each individuals with trait values in the interval (x_1, x_2) are given a particular rating (A) by a judge.

The scale value corresponding to this rating A is defined to be the mean trait value of all these individuals and is accordingly given by the formula (due to Likert) :

$$\text{Scale value} = \int_{x_1}^{x_2} x f(x|A)\, dx$$ where, $f(x|A)$ is the truncated p.d.f. of X over the interval $A = (x_1, x_2)$ and

$$f(x|A) = \begin{cases} \dfrac{\phi(x)}{P(A)}, & \text{if } x \in A \\ 0, & \text{OW} \end{cases}$$

Hence, scale value $= \displaystyle\int_{x_1}^{x_2} x \cdot \frac{\phi(x)}{P(A)}\, dx$

$$= \frac{\int_{x_1}^{x_2} x\phi(x)\, dx}{\Phi(x_2) - \Phi(x_1)} = \frac{\left[-\phi(x)\right]_{x_1}^{x_2}}{\Phi(x_2) - \Phi(x_1)}$$

$$= \frac{\phi(x_1) - \phi(x_2)}{\Phi(x_2) - \Phi(x_1)}.$$

After obtaining the scale values for different ratings of different judges, different judges may be compared. The scale is known as Likert's scale.

Scanned by CamScanner

(b) Scaling of Rankings :– Suppose N individuals are ranked by a judge in order of merit of a particular trait. Under the assumption of the normality of the trait, gives that a rank R of an individual represents the interval from $R-\frac{1}{2}$ to $R+\frac{1}{2}$. Hence, in case of no tie, cumulative frequency of greater than type of the rank R is $(N-R+0\cdot5)$. Then the percentile score corresponding to rank R of an individual among N individuals is given by :

$$p_R = 100 \cdot \frac{N-R+\frac{1}{2}}{N} = 100\left\{1-\frac{R-\frac{1}{2}}{N}\right\}$$

This formula enables us to convert any set of ranks into scores if we are justified in assuming normality in the trait for which ranks are given. The scale values corresponding to p_R's are obtained by finding normalised scores ξ's corresponding to p_R by the relation:

$$\frac{p_R}{100} = \int_{-\infty}^{\xi_i} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(\xi).$$

EXAMPLE :–

1. Assuming the trait density to be $f(x)=\frac{1}{2}e^{-|x|}, x\in\mathbb{R}$, describe Likert's method of scaling ratings. Obtain explicit expressions of the scale-values for a class.

Solution:-

Here, the trait$(x) \sim$ Double exponential distn.

The scale-value for a class (x_1, x_2) is

$$\frac{\int_{x_1}^{x_2} x f(x)dx}{\int_{x_1}^{x_2} f(x)dx} = \frac{\int_{x_1}^{x_2} x\cdot\frac{1}{2}e^{-|x|}dx}{\int_{x_1}^{x_2}\frac{1}{2}e^{-|x|}dx}.$$

Case-I: If $x_2 > x_1 > 0$, then scale value $= \dfrac{\int_{x_1}^{x_2} x \cdot \frac{1}{2} e^{-x}\,dx}{\int_{x_1}^{x_2} \frac{1}{2} e^{-x}\,dx}$

$$= \frac{\left[-xe^{-x} - e^{-x}\right]_{x_1}^{x_2}}{e^{-x_1} - e^{-x_2}}$$

$$= 1 + \frac{x_1 e^{-x_1} - x_2 e^{-x_2}}{e^{-x_1} - e^{-x_2}}.$$

Case-II: If $x_1 < 0 < x_2$, then the scale-value is

$$= \frac{\int_{x_1}^{0} x \cdot \frac{1}{2} e^{x}\,dx + \int_{0}^{x_2} x \cdot \frac{1}{2} e^{-x}\,dx}{\int_{x_1}^{0} \frac{1}{2} e^{x}\,dx + \int_{0}^{x_2} \frac{1}{2} e^{-x}\,dx}$$

$$= \frac{\left[xe^{x} - e^{x}\right]_{x_1}^{0} + \left[-e^{-x}(x+1)\right]_{0}^{x_2}}{\left[e^{x}\right]_{x_1}^{0} + \left[-e^{-x}\right]_{0}^{x_2}}$$

$$= \frac{(1-x_1)e^{x_1} - (1+x_2)e^{-x}}{2 - (e^{x_1} + e^{-x_2})}$$

Case-III: If $x_1 < x_2 < 0$, then the scale-value is

$$= \frac{\int_{x_1}^{x_2} \frac{x}{2} e^{x}\,dx}{\int_{x_1}^{x_2} \frac{1}{2} e^{x}\,dx}$$

$$= \frac{\left[(x-1)e^{x}\right]_{x_1}^{x_2}}{\left[e^{x}\right]_{x_1}^{x_2}}$$

$$= \frac{(x_2 e^{x_2} - x_1 e^{x_1})}{e^{x_2} - e^{x_1}} - 1.$$

Limit Theorems

LIMIT THEOREMS

LAW OF LARGE-NUMBERS :→ In practice estimates are made of an unknown quantity (Parameter) by taking the average ($\overline{X}_n$) of a number of repeated measurements of the quantity each of which may be in error. It is, therefore, of interest to study the properties of such an estimate. An initial enquiry is made concerning its behaviour as the number of measurements (n) increases ($\to \infty$). Does the estimate converge in some sense to the true value (ξ) of the parameter under study?

The probability can be formulated in the following way. Let $\{X_n\}$ be a sequence of observation and $\overline{X_n}$ be the average of the first n observations. Under what condition can we use that, $\overline{X_n} \to \xi$ (the unknown quantity)

In one or other sense we shall generalise the problem further and ask for the conditions under which $(\overline{X_n} - \overline{\xi}_n) \to 0$ where $\{\xi_n\}$ is a sequence of constants is saught to be measured by sequence of observations $\{X_n\}$. We shall say that law of large number holds if the convergence such as,

$$\overline{X_n} \to \xi \text{ or, } (\overline{X}_n - \xi) \to 0 \text{ takes place.}$$

When the convergence is in probability we shall say that Weak Law of Large numbers (WLLN) holds. Thus the theorems on WLLN states the conditions under which the WLLN holds for a sequence of random variables $\{X_n\}$.

In other words, our problem is to answer the question in the affirmative sense that whether $\exists$ sequence of constants $\{A_n\}$ and $\{B_n\}$ $(B_n > 0)$, $B_n \to \infty$ as $n \to \infty$ such that,

$$\frac{S_n - A_n}{B_n} \to 0 \text{ as } n \to \infty, \text{ where } S_n = \sum_{K=1}^{n} X_K,\ n=1,2,\ldots$$

Weak Law of Large Numbers :—

- Definition :→ Let $\{X_n\}$ be a sequence of random variables and let $S_n = \sum_{K=1}^{n} X_K$, $n=1,2,\ldots\ldots$ We say that $\{X_n\}$ obey WLLN with respect to the sequence of constants $\{B_n\}$, $B_n > 0$, $B_n \uparrow \infty$ as $n \to \infty$, if there exists a sequence of real constants A_n such that,

$$P\left\{ \left| \frac{S_n - A_n}{B_n} \right| > \epsilon \right\} \to 0 \text{ as } n \to \infty.$$

A_n are called centering constants, and B_n are called Norming constants.

• Definition 2. :→ Let $\{X_n\}$ be a sequence of random variable. Define, $S_n = \sum_{K=1}^{n} X_K$, suppose $\{B_n\}$ is a sequence of random variable with $E(X_K) = \mu_K$. Define, $\overline{X}_n = \frac{1}{n}\sum_{K=1}^{n} X_K$, $\overline{\mu}_n = \frac{1}{n}\sum_{K=1}^{n}\mu_K$, the sequence obeys WLLN if given $\epsilon > 0$, $\delta > 0$, we can choose a natural number $K = K(\epsilon, \delta) \ni$ whenever $n \geq K$,

$$P[|\overline{X}_n - \overline{\mu}_n| < \epsilon] > 1 - \delta$$

Provided $V(\overline{X}_n) \to 0$

Note:- The condition $V(\overline{X}_n) \to 0$ as $n \to \infty$ is a sufficient condition for the sequence of random variable $\{X_n\}$ in order to obey WLLN. But the converse may not be true, i.e. in case of $V(\overline{X}_n) \not\to 0$, we have no conclusion in that case.

WLLN (Weak Law of Large Numbers)

Statement: — With the probability approaching unity or certainty as near as we please, we may expect that the arithmatic mean of values actually assumed by the random variables will differ from the arithmatic mean of their expectations by less than a given numbers, however small, provided the number of variables can be taken sufficiently large and provided the variance of the arithmatic mean of the random variables approaches to zero as $n \to \infty$.

☆ Markov's Theorem (WLLN) :— If $\{X_n\}$ be a sequence of random variables with $\mu_1, \mu_2, \ldots$ their expectations and B_n = Variance $\left(\sum_{K=1}^{n} X_K\right) < \infty$ then,

$$P\{|\overline{X}_n - \overline{\mu}_n| \geq \epsilon\} \to 0 \text{ as } n \to \infty$$

if $\frac{B_n}{n^2} \to 0$ as $n \to \infty$, where, $\overline{X}_n = \frac{1}{n}\sum_{K=1}^{n} X_K$ and

$$\overline{\mu}_n = \frac{1}{n}\sum_{K=1}^{n} \mu_K .$$

Summary: — ① $E(X_K) = \mu_K < \infty$

② $V\left(\sum_{K=1}^{n} X_K\right) < \infty$

③ $\frac{B_n}{n^2} \to 0$ as $n \to \infty$

WLLN holds for $\{X_n\}$ w.r.t. $\{\mu_n\}$, $\{n\}$

i.e. $P[|\overline{X}_n - \overline{\mu}_n| \geq \epsilon] \to 0$ as $n \to \infty$

Proof:— $P\{|\bar{X}_n-\bar{\mu}_n|\geq \epsilon\}$

$$= P\{(\bar{X}_n-\bar{\mu}_n)^2 \geq \epsilon^2\} \leq \frac{E(\bar{X}_n-\bar{\mu}_n)^2}{\epsilon^2}$$ [By Markov's inequality]

Now, $E(\bar{X}_n-\bar{\mu}_n)$

$$= \frac{1}{n}\sum_{k=1}^{n}\{E(X_k)-\mu_k\} = 0 \quad \because E(X_k)=\mu_k \ \forall k$$

and; $E(\bar{X}_n-\bar{\mu}_n)^2 = Var(\bar{X}_n)_k$

$$= Var\left(\frac{1}{n}\sum_{1=1} X_k\right)$$

$$= \frac{1}{n^2} Var\left(\sum_{1=1}^{k} X_k\right)$$

$$= \frac{B_n}{n^2}.$$

As assumed by the condition of the theorem, we have —

$$\frac{B_n}{n^2} \to 0 \text{ as } n\to\infty$$

i.e. $E(\bar{X}_n-\bar{\mu}_n)^2 \to 0$ as $n\to\infty$

$$\therefore 0 \leq P[|\bar{X}_n-\bar{\mu}_n|\geq\epsilon] \leq \frac{E(\bar{X}_n-\bar{\mu}_n)^2}{\epsilon^2} \to 0 \text{ as } n\to\infty$$

i.e. $P[|\bar{X}_n-\bar{\mu}_n|\geq\epsilon] \to 0$ as $n\to\infty$.

☆ **Theorem:**– If the variables $\{X_n\}$ are uniformly bounded then the condition $\frac{B_n}{n^2}\to 0$ as $n\to\infty$ is necessary and as well as sufficient.

Proof:→ Let $\xi_k = X_k-\mu_k \ \forall k=1,2,\dots$

Since, X_k's are uniformly bounded.

$\therefore \exists$ a positive number $c \ni$

$|\xi_k| > c \quad \forall k=1,2,\dots$

Let, $P = P\{|\xi_1+\xi_2+\dots+\xi_n|\leq n\epsilon\}$

$$\Rightarrow 1-P = P\{|\xi_1+\xi_2+\dots+\xi_n| > n\epsilon\}$$

$$= P\{|U_n| > n\epsilon\} \quad \text{(say)}$$

then $E(U_n)=0$, $\therefore Var(U_n) = E(U_n^2) = B_n$ (say)

$$\therefore B_n = \int_0^\infty U_n^2\, dF(U_n) = \int_0^{n\epsilon} u_n^2\, dF(u_n) + \int_{n\epsilon}^\infty u_n^2\, dF(u_n)$$

$$\leq n^2\epsilon^2 P + n^2\epsilon^2(1-P)$$

Since, $|u_n| = |\xi_1+\xi_2+\dots+\xi_n| \leq \left\{\sum_{k=1}^{n}|\xi_k|\right\} < nc$

i.e. $\frac{B_n}{n^2} \leq \epsilon^2 P + c^2(1-P)$

< arbitrary small positive quantity if $1-P\to 0$
(as X_n's are bounded)

$\therefore X_n$'s are bounded $\Rightarrow \frac{B_n}{n^2}\to 0$ as $n\to\infty$.

☆ **Chebyshev's Theorem (WLLN):** — Let $\{X_K\}$ be a sequence of independent random variables such that $E(X_K) = \mu_K$ and $V(X_K) = \sigma_K^2$ exists $\forall\ K = 1, 2, \ldots$

Then $P[|\bar{X}_n - \bar{\mu}_n| \leq \epsilon] \to 1$ as $n \to \infty$

WLLN holds for the sequence $\{X_K\}$ if $\lim_{n\to\infty} \frac{1}{n^2}\sum_{K=1}^{n} \sigma_K^2 = 0$.

Proof: — Here, $E\{\bar{X}_n - \bar{\mu}_n\} = 0$

$$\text{var}\{\bar{X}_n - \bar{\mu}_n\} = E\{\bar{X}_n - \bar{\mu}_n\}^2 = \text{var}(\bar{X}_n) = \text{Var}\left\{\frac{1}{n}\sum_{K=1}^{n} X_K\right\}$$

$$= \frac{1}{n^2}\sum_{K=1}^{n} \sigma_K^2,$$

since X_K's are independent

$$P[|\bar{X}_n - \bar{\mu}_n| \leq \epsilon] = P[(\bar{X}_n - \bar{\mu}_n)^2 \leq \epsilon^2] \text{ for } \epsilon > 0$$

$$= P\left[(\bar{X}_n - \bar{\mu}_n)^2 \leq \left(\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2\right) \cdot \frac{\epsilon^2}{\left(\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2\right)}\right]$$

$$\geq 1 - \frac{\sum_{K=1}^{n}\sigma_K^2}{n^2\epsilon^2} \quad \{\text{by chebyshev's inequality}\}$$

$$\to 1 \text{ as } n \to \infty$$

since, $\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2 \to 0$ as $n \to \infty$.

Alternative Proof (Cor. WLLN): —

$$B_n = \text{var}\left(\sum_{K=1}^{n} X_K\right) = \sum_{K=1}^{n}\text{var}(X_K) = \sum_{K=1}^{n}\sigma_K^2, \quad \because X_K\text{'s are independent.}$$

$$\therefore \frac{B_n}{n^2} = \frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2.$$

By WLLN, $\frac{B_n}{n^2} \to 0$ as $n \to \infty$ implies that WLLN holds for $\{X_K\}$

i.e. $\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2 \to 0$ as $n\to\infty$ implies,

$$P[|\bar{X}_n - \bar{\mu}_n| \leq \epsilon] \to 1 \text{ as } n \to \infty.$$

▣ **Cor. to Chebyshev's theorem (WLLN):** — If $\sigma_K^2 = \sigma^2\ \forall\ K = 1, 2, \ldots, n$

i.e. X_K's are homoscedastic apart from being independent then WLLN holds for $\{X_K\}$.

This is because $\lim_{n\to\infty}\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2 = \lim_{n\to\infty}\frac{n\sigma^2}{n^2} = 0$ (always)

i.e. in that case the criterion $\lim_{n\to\infty}\frac{1}{n^2}\sum_{K=1}^{n}\sigma_K^2 = 0$ is obvious, i.e. X_K's are independent.

$\text{var}(X_K) = \sigma^2 < \infty\ \forall\ K = 1, 2, \ldots$

$\Rightarrow$ WLLN holds for $\{X_K\}$

i.e. $P[|\bar{X}_n - \bar{\mu}_n| \leq \epsilon] \to 1$ as $n \to \infty$

☆ **Kinchin's Theorem (WLLN)** :— Let $\{X_K\}$ be a sequence of i.i.d. random variables with finite expectation $E(X_K)=\mu$. Then $P[|\bar{X}_n-\mu|\le\epsilon]\to 1$ as $n\to\infty$, i.e. WLLN holds for $\{X_K\}$.

Proof:— Let $\phi(t)$ be the common characteristic function of X_K. Then the c.f. of $\bar{X}_n$ is,

$$\phi_n(t)=E\{e^{it\bar{X}_n}\}$$
$$=E\left\{e^{it\cdot\frac{1}{n}\sum_{K=1}^{n}X_K}\right\}$$
$$=\prod_{K=1}^{n}\left(e^{i\cdot t/n\cdot X_K}\right) \quad \because X_K\text{'s are independent.}$$
$$=\left\{\phi\left(\frac{t}{n}\right)\right\}^n \quad \because X_K\text{'s are identical.}$$

Now, where $E(X_K)$ exists then $\phi(t)$ can be written as

$$\phi(t)=1+it\mu+O(t), \text{ where, } O(t) \text{ is } \ni \frac{O(t)}{t}\to 0 \text{ as } t\to 0$$
$$\therefore \phi_n(t)=\left\{1+i\frac{t}{n}\mu+O\left(\frac{t}{n}\right)\right\}^n$$
$$\log\phi_n(t)=n\log\left(1+i\cdot\frac{t}{n}\mu+O\left(\frac{t}{n}\right)\right)$$
$$=n\left[i\cdot\frac{t}{n}\mu+O\left(\frac{t}{n}\right)\right]$$
$$=it\mu+n\cdot O\left(\frac{t}{n}\right).$$
$$\longrightarrow it\mu \text{ as } n\to\infty$$

since $\lim_{n\to\infty} n\cdot O\left(\frac{t}{n}\right) = t\left\{\lim_{\frac{t}{n}\to 0}\frac{O\left(\frac{t}{n}\right)}{\left(\frac{t}{n}\right)}\right\}=t\times 0=0$

i.e. $\phi_n(t)\to e^{it\mu}$ as $n\to\infty$

But $e^{it\mu}$ is the c.f. of a distribution degenerate at μ.

i.e. $P[|\bar{X}_n-\mu|\le\epsilon]\to 1$ as $n\to\infty$

WLLN holds for i.i.d. random variables $\{X_K\}$ with finite mean.

☆ **Bernoulli's Theorem (WLLN)** :— If $\{X_K\}$ be a sequence of (independent) Bernoulli's random variables with common probability p then,

$$P[|\bar{X}_n-p|\le\epsilon]\to 1 \text{ as } n\to\infty$$
$$P\left[\left|\frac{S_n}{n}-p\right|\le\epsilon\right]=P[|S_n-np|\le n\epsilon]\to 1 \text{ as } n\to\infty.$$

Proof:— since X_K's are Bernoulli random variables,

$$X_K=\begin{cases}1 & \text{with probability } p\\ 0 & \text{with probability } (1-p).\end{cases}$$

$$E(X_K)=p,\quad E(X_K^2)=p,\quad Var(X_K)=pq\ \forall\ K=1,2,\ldots$$
$$\therefore Var(X_K)=pq\le\frac{1}{4}.$$

$$\therefore P\left[\left|\frac{S_n}{n}-p\right| \le \epsilon\right] = P\left[\left(\frac{S_n}{n}-p\right)^2 \le \epsilon^2\right]$$

$$= P\left[(S_n-np)^2 \le n^2\epsilon^2\right]$$

Now, $S_n = n\bar{X}_n = \sum_{k=1}^{n} X_k \sim Bin(n,p)$

$E(S_n) = np$ and $V(S_n) = npq$.

$\therefore E(S_n-np)^2 = npq$

$$\therefore P\left[\left|\frac{S_n}{n}-p\right| \le \epsilon\right] = P\left[(S_n-np)^2 \le npq \cdot \frac{n^2\epsilon^2}{npq}\right] \geq 1-\frac{npq}{n^2\epsilon^2}$$

$$= 1-\frac{pq}{n\epsilon^2} \quad \text{[by Chebyshev's inequality]}$$

$\to 1$ as $n\to\infty$

$\because pq \le \frac{1}{4}$ and $\epsilon>0$ is small.

i.e. $P\left[\left|\frac{S_n}{n}-p\right| \le \epsilon\right] \to 1$ as $n\to\infty$

or, $P\left[|\bar{X}_n - p| \le \epsilon\right] \to 1$ as $n\to\infty$

☒ Cor. to Chebyshev's theorem: —

$\sigma_k^2 = V(X_k) = pq \le \frac{1}{4} \ \forall k$

By Chebyshev's theorem (WLLN)

$P\left[|\bar{X}_n - p| \le \epsilon\right] \to 1$ as $n\to\infty$

☒ Cor. to Kinchin's theorem: —

$E(X_k) = p$ and X_k's are i.i.d.

By Kinchin's theorem (WLLN)

$\bar{X}_n \xrightarrow{P} \mu = p$

i.e. $P\left[|\bar{X}_n - p| \le \epsilon\right] \to 1$ as $n\to\infty$

⇨ Concept of O and o: —

$f(x) = O\{g(x)\} \Rightarrow f(x)$ is at most of the order $g(x)$ as $x\to\infty$

if $\lim_{x\to\infty} \frac{f(x)}{g(x)} =$ constant, then we write

$f(x) = O\{g(x)\}$

Similarly, if $\lim_{x\to 0} \frac{f(x)}{g(x)} = 0$ then we write $f(x) = o\{g(x)\}$

and we say that $f(x)$ is of smallest order than $g(x)$

lemma:— Let us write $f(x) = o(x)$ if $\frac{f(x)}{x} \to 0$ as $x \to 0$

We have $\lim_{n\to\infty} \left\{1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right\}^n = e^a$ for every real a.

By Taylor's theorem, we have

$$f(x) = f(0) + x f'(\theta x) \quad , \quad 0 < \theta < 1$$

$$\Rightarrow f(x) = f(0) + x f'(0) + x\left\{f'(\theta x) - f'(0)\right\}$$

if $f'(x)$ is continuous at $x=0$ then,

$$f(x) = f(0) + x f'(0) + o(x)$$

taking $f(x) = \log(1+x)$

$\therefore f'(x) = \frac{1}{1+x}$ which is continuous at $x=0$

$$\therefore \log(1+x) = \log 1 + x + o(x)$$

i.e. $\log(1+x) = x + o(x)$

then for sufficiently large n,

$$n \log\left\{1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right\} = n\left[\frac{a}{n} + o\left(\frac{1}{n}\right) + o\left\{\frac{a}{n} + o\left(\frac{1}{n}\right)\right\}\right]$$

$$= a + n\,o\left(\frac{1}{n}\right) + n\,o\left\{\frac{a}{n} + o\left(\frac{1}{n}\right)\right\}$$

$$= a + n\,o\left(\frac{1}{n}\right) + n\,o\left(\frac{1}{n}\right)$$

$$\to a \quad \text{as } n \to \infty$$

i.e. $\lim_{n\to\infty} \left\{1 + \frac{a}{n} + o\left(\frac{1}{n}\right)\right\}^n = e^a$, for every real a.

Note:— $o\left(\frac{1}{n}\right) = \frac{k_1}{n^2} + \frac{k_2}{n^3} + \cdots\cdots$

$$o\left(\frac{a}{n} + o\left(\frac{1}{n}\right)\right) = k_1\left(\frac{a}{n} + o\left(\frac{1}{n}\right)\right)^2 + k_2\left(\frac{a}{n} + o\left(\frac{1}{n}\right)\right)^2 + \cdots\cdots$$

$$= \frac{c_1}{n^2} + \frac{c_2}{n^3} + \cdots\cdots \quad \text{(say)}$$

$$= o\left(\frac{1}{n}\right).$$

Convergence in Distribution: —

- **Definition:** — Let $\{F_n\}$ be a sequence of distribution function and F be another distribution function. $\{F_n\}$ converges in distribution or, weakly converges to F if $F_n(x) \to F(x)$ whenever $n \to \infty$ at all continuity points x of $F(x)$.
Symbolically, we denote, → $F_n \xrightarrow{\omega} F$

Let $\{X_n\}$ be a sequence of random variables with corresponding sequence of distribution function $\{F_n\}$ and X be another random variable with distribution function F.

X_n converges in Law or, in distribution to X if $F_n \xrightarrow{\omega} F$.
We denote, $X_n \xrightarrow{L} X$

or, $X_n \xrightarrow{D} X$

EXAMPLES: →

① It may happen that $F_n \xrightarrow{\omega} F$, where F is not at all a distribution function.

Define, $F_n(x) = \begin{cases} 0 & \text{if } x < n \\ 1 & \text{if } x \geq n \end{cases}$, $n = 1, 2, \ldots$

As $n \to \infty$, $F_n(x) \to F(x) = 0 \ \forall x \in \mathbb{R}'$
but $F(x) = 0$, $x \in \mathbb{R}'$ is surely not a distribution function.

② Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from rectangular $(0, \theta)$ Population.

Define, $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$

note that,
$$P[X_{(n)} \leq x] = F_n(x)$$
$$= P[X_1, X_2, \ldots, X_n \leq x]$$
$$= \prod_{i=1}^{n} P[X_i \leq x]$$
$$= [P[X_1 \leq x]]^n \quad [\because X_i\text{'s are identically distributed}]$$
$$= \begin{cases} 0 & \text{if } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{if } 0 < x < \theta \\ 1 & \text{if } x \geq \theta \end{cases}$$

Note that, $F_n(x) \xrightarrow{\omega} F(x) = \begin{cases} 0 & \text{if } x < \theta \\ 1 & \text{if } x \geq \theta \end{cases}$

Thus we see that the limiting distribution become degenerate at θ.

③ Let X_n be a random variable having the distribution function F_n,

$$F_n(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \frac{1}{n} & \text{if } 0 \le x < n \\ 1 & \text{if } x \geq n \end{cases}$$

$$X_n \xrightarrow{L} X$$

if F be the distribution function of X then F is given by,

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

i.e. X is degenerate at 0, but the convergence in distribution may not imply the convergence in moments.

Note that, — $E(X_n^K) = n^K \cdot \frac{1}{n} = n^{K-1}$

$$E(X^K) = 0$$

$$E(X_n^K) \not\rightarrow E(X^K)$$

It may happen that the limiting function of a probability mass function f_n is not at all a p.m.f.

$$f_n(x) = P[X_n = x] = \begin{cases} 1 & \text{if } x = 2 + \frac{1}{n} \\ 0 & \text{o/w} \end{cases}$$

Note that, none of the f_n's assign any mass to the point 2.
where $f_n(x) \rightarrow f(x)$ whenever $n \to \infty$ with $f(x) = 0 \ \forall x \in \mathbb{R}$ though the limiting distribution becomes degenerate at 2.

☒ <u>Convergence in Probability</u> : — Let $\{X_n\}$ be a sequence of RVs defined on some probability space $(\Omega, \mathcal{A}, P)$. We say that the sequence $\{X_n\}$ converges in probability to the RV X if for every $\epsilon > 0$,

$$P\{|X_n - X| > \epsilon\} \rightarrow 0 \quad \text{as } n \to \infty$$

or, equivalently,

$$P\{|X_n - X| < \epsilon\} \rightarrow 1 \quad \text{as } n \to \infty$$

We write, $X_n \xrightarrow{P} X$.

<u>Remark</u>: — We emphasize that the definition says nothing about the convergence of the RVs X_n to the RV X in the sense in which it is understood in real analysis.
Thus $X_n \xrightarrow{P} X$ does not imply that given $\epsilon > 0$, we can find an N such that $|X_n - X| < \epsilon$ for $n \geq N$. Above definition speaks about the convergence of the sequence of probabilities $P\{|X_n - X| > \epsilon\}$ to 0.

[C.U. 2005]

Example: Let $\{X_n\}$ be a sequence of RVs with PMF
$P\{X_n = 1\} = \frac{1}{n}$, and $P\{X_n = 0\} = 1 - \frac{1}{n}$.

Then $$P\{|X_n| > \epsilon\} = \begin{cases} P\{X_n = 1\} = \frac{1}{n} & \text{if } 0 < \epsilon < 1 \\ 0 & \text{if } \epsilon \geq 1. \end{cases}$$

It follows that $P\{|X_n| > \epsilon\} \to 0$ as $n \to \infty$, and we conclude that $X_n \xrightarrow{P} 0$.

Result-1.

i) $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$ but the converse may not be true.

ii) $X_n \xrightarrow{P} C$ [Constant] $\Leftrightarrow X_n \xrightarrow{L} C$

Result-2.

i) (a) $X_n \xrightarrow{L} X$, g is a continuous function,
$\Rightarrow g(X_n) \xrightarrow{L} g(X)$

(b) $X_n \xrightarrow{L} C$ [Constant]
g is continuous at the neighbourhood of C.
$\Rightarrow g(X_n) \xrightarrow{L} g(C)$

ii) (a) and (b), for convergence in probability we get the same fact.

Result-3. Suppose $X_n \xrightarrow{P} C$ [Constant]
$Y_n \xrightarrow{L} Y$

$\Rightarrow$ i) $X_n \pm Y_n \xrightarrow{L} C \pm Y$

ii) $X_n Y_n \xrightarrow{L} CY$

in particular if $C = 0$
$X_n Y_n \xrightarrow{L} 0$

iii) $\frac{X_n}{Y_n} \xrightarrow{L} \frac{Y}{C}$ (Provided $C \neq 0$)

Result-4. i) $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, a, b constants $\Rightarrow X_n Y_n \xrightarrow{P} ab$, for
$$X_n Y_n = \frac{(X_n + Y_n)^2 - (X_n - Y_n)^2}{4} \xrightarrow{P} \frac{(a+b)^2 - (a-b)^2}{4} = ab.$$

ii) $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, a, b constants, $b \neq 0$
$\Rightarrow X_n Y_n^{-1} \xrightarrow{P} ab^{-1}$.

Remark:— We emphasize that we can not improve the result above by replacing k by an RV; i.e. $X_n \xrightarrow{L} X$, in general does not imply $X_n \xrightarrow{P} X$, for let $X, X_1, X_2, \ldots$ be identically distributed RVs and let the joint distribution of (X_n, X) be as follows:

X \ X_n	0	1	
0	0	1/2	1/2
1	1/2	0	1/2
	1/2	1/2	1

Clearly, $X_n \xrightarrow{L} X$, but

$$P\{|X_n - X| > \tfrac{1}{2}\} = P\{|X_n - X| = 1\}$$
$$= P\{X_n = 0, X = 1\} + P\{X_n = 1, X = 0\}$$
$$= 1 \not\to 0$$

Hence, $X_n \overset{P}{\not\to} X$ but $X_n \xrightarrow{L} X$

Normal Convergence:— If $F_n \xrightarrow{w} F$, where F is the DF of a normal distribution, then this convergence is called a normal convergence and we say that central limit law holds for the sequence of random variables $\{X_n\}$ having $\{F_n\}$ as their DFs.

• **Definition:**— Let the distribution of a random variable Y depends on a parameter n and there exists two quantities μ and σ^2 (which may or may not depend on n) such that,

$$P_n\left[\frac{Y-\mu}{\sigma} \le t\right] \longrightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-x^2/2}\,dx = \Phi(t) \quad \text{as } n \to \infty.$$

Then we say that Y is asymptotically normally distributed with mean μ and variance σ^2 or, that $\frac{Y-\mu}{\sigma}$ follows the central limit law or normal convergence. The quantity μ and σ^2 are called the asymptotic mean and asymptotic variance, respectively.

$$\mu = \text{a mean}(Y)$$
$$\sigma^2 = \text{a variance}(Y)$$

Note:→ $\left.\begin{array}{l}\mu = \text{a mean}(Y) \\ \sigma^2 = \text{a variance}(Y)\end{array}\right\} \not\Rightarrow \begin{array}{l} E_n(Y) \to \mu \\ V_n(Y) \to \sigma^2 \quad \text{as } n \to \infty.\end{array}$

Scanned by CamScanner

Central Limit Theorem :—

Let $\{X_n\}$ be a sequence of independent random variable and $\{A_n\}$, $\{B_n\}$ be sequences of contering constant and norming constant. Define,

$$S_n = \sum_{K=1}^{n} X_K$$

Then under very general condition if

$$P\left(\frac{S_n - A_n}{B_n} \le x\right) \to \Phi(x) \text{ as } n \uparrow \infty .$$

$\Phi(x)$ being distribution function of $N(0,1)$, we say that CLT holds good.

Remarks:— If $\{X_n\}$ be a sequence of independent random variable, with

$$E(X_n) = \mu_n \text{ and } V(X_n) = \sigma_n^2$$

Then our choice of $\{A_n\}$ and $\{B_n\}$ would be respectively,

$$A_n = \sum_{K=1}^{n} \mu_K \quad \text{and } B_n = \sqrt{\sum_{K=1}^{n} \sigma_K^2} .$$

If, further, $\{X_n\}$ be a sequence of i.i.d. random variable

$$E(X_n) = \mu \text{ and } V(X_n) = \sigma^2$$

then we may choose $A_n = n\mu$ and $B_n = \sigma\sqrt{n}$.

Remarks:— Quantile – Quantile Plot

In order to ensure, the asymptotic normality empirically for moderate sample sizes. A useful device is Q–Q plot.

Which is the library function in most of the statistical packages.

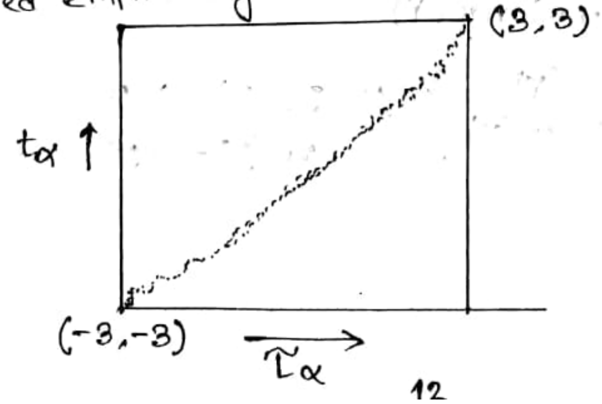
Let, t_α be the $(1-\alpha)$th quantile or fractile of $\frac{S_n - A_n}{B_n}$, then CLT states that,

$$t_\alpha \to \tau_\alpha \quad \forall \alpha \in (0,1)$$

τ_α being the upper α point of $N(0,1)$ distribution.

i.e. the $(1-\alpha)$th quantile of the distribution.

If from the quantile - quantile plot, normal distribution is justified empirically, then the plot looks like



Remarks:—

1. Here it may happen that $\{X_n\}$ be a sequence of discrete random variable but the asymptotic distribution of $\left\{\frac{S_n - A_n}{B_n}\right\}$ is absolutely continuous as it is normal.

2. Asymptotic normality can also be ensured for a sequence of dependent random variables $\{X_n\}$ [in this regard, there are some results due to Hajek and Sidak].

3. Loeve showed that asymptotic distribution of $\frac{S_n - A_n}{B_n}$ has to be a stable distribution,

* If X_1, X_2 be i.i.d. non-degenerate RV with common distribution function F, which is said to be stable provided the distribution of $\frac{a_1 X_1 + a_2 X_2 - A}{B}$ has the DF F, $\forall$ $a_1, a_2 > 0$ and the centering constant A, norming constant B are to be determined on the basis of (a_1, a_2). *

Domain of Attraction:— Let $X_1, X_2, \ldots$ be i.i.d. RVs with common distribution function F. We say that F belongs to the domain of attraction of a distribution V, if $\exists$ norming constants $B_n > 0$ and centering constant A_n such that as $n \to \infty$

$$P\{B_n^{-1}(S_n - A_n) \le x\} \to V(x)$$

at all continuity points x of V if V become Φ, we get CLT.

CLT (Central limit Theorem):— (Another approach)

Let $\{X_n\}$ be a sequence of RVs such that under certain given condition the distribution of

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \text{ i.e. of } \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$$

is asymptotically normal with mean 0 and variance 1, then we say that central limit law holds for $\{X_n\}$.
i.e. $\{X_n\}$ obeys central-limit-law.
A central limit theorem provides those conditions under which it holds.

Continuity theorem for C.F.'s :→ For sequence of distribution functions $\{F_n\}$ with C.F.s $\{\phi_n\}$, a necessary and sufficient condition that $F_n(x) \xrightarrow{w} F(x)$ at all continuity points of F is that for every real t $\phi_n(t) \longrightarrow \phi(t)$ which is continuous at $t = 0$ and $\phi(t)$ is the CF of F.

Now, $\phi(t) = e^{-t^2/2}$ is the CF of N(0,1) distribution. So, if for a sequence of random varibles $\{X_n\}$ we have $\phi_n(t) \to e^{-t^2/2}$ as $n \to \infty$, Then we say that Central Limit law holds for $\{X_n\}$ where,

$$\phi_n(t) = \text{c.f. } \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}.$$

Lindeberg - Lévy Limit Theorem (CLT): —

- **Statement:** Let $\{X_n\}$ be a sequence of i.i.d. random variables, with $E(X_n) = \mu$ and $V(X_n) = \sigma^2 < \infty$.

Define, $S_n = \sum_{K=1}^{n} X_K$;

Then. $P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x \right) \to \Phi(x)$ as $n \uparrow \infty$, where $\Phi(x)$ = distribution function of $N(0,1)$.

$\therefore \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim AN(0,1)$, i.e. $\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{L} X \sim N(0,1)$.

<u>Proof</u>:- Define, $Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{K=1}^{n} \frac{(X_K - \mu)}{\sigma}$

$$= \frac{1}{\sqrt{n}} \sum_{K=1}^{n} Z_K .$$

Characteristic function of Z is, —

$$\phi_Z(t) = E\left(e^{itZ}\right)$$

$$= E\left(e^{it \cdot \frac{1}{\sqrt{n}} \sum_{K=1}^{n} Z_K}\right)$$

$$= E\left(\prod_{K=1}^{n} e^{\frac{itZ_K}{\sqrt{n}}}\right)$$

$$= \prod_{K=1}^{n} E\left(e^{itZ_K/\sqrt{n}}\right)$$ [due to independence] no relationship between the random variables.

$$= \prod_{K=1}^{n} \phi_{Z_K}\left(\frac{t}{\sqrt{n}}\right)$$

$$= \left[\phi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$ [as the sequence of R.V.s are i.i.d.]

$$= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$ [$\because E(Z_1) = 0$ and $E(Z_1^2) = 1$]

[as $\phi(t) = 1 - \frac{1}{2}t^2 + t^2 o(1)$] $\therefore \phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n}t^2 + \frac{t^2}{n} o(1)$.

Now, $\lim_{n\to\infty} \phi_Z(t) = \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n$

$$= \lim_{n\to\infty} \left(1 - \frac{t^2}{2n}\right)^n$$ [as, $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$]

$$= e^{-t^2/2}$$, which is the CF of a $N(0,1)$.

By continuity theorem of the CF,

$Z \xrightarrow{L} N(0,1)$. i.e. CLT holds for $\{X_n\}$.

De-Moivre Laplace Limit theorem:—

- **Statement:**— Let $\{X_n\}$ be a sequence of independent Bernoulli random variable with common probability p,

i.e. $P\{X_n=1\} = p = 1-P\{X_n=0\}$.

Define $S_n = \sum_{K=1}^{n} X_K$, then

$$P\left[\frac{S_n - np}{\sqrt{npq}} \le x\right] \to \Phi(x) \text{ as } n\to\infty \text{ \& } q=1-p.,$$

i.e. $\frac{S_n - np}{\sqrt{npq}} \sim AN(0,1)$ or, $S_n \xrightarrow{L} X \sim N(np, npq)$;

where, $\frac{S_n-np}{\sqrt{npq}}$ is asymptotically normal with mean zero and unit variance.

Proof:— $S_n = \sum_{K=1}^{n} X_K$ and $E(S_n) = np$, $V(S_n) = \sqrt{npq}$

then, $\phi_n(t) = E\left[\exp\left\{\left(\frac{S_n-np}{\sqrt{npq}}\right)it\right\}\right]$

$$= \prod_{K=1}^{n} E\left[\exp\left\{\frac{it}{\sqrt{n}}\left(\frac{X_K-p}{\sqrt{p(1-p)}}\right)\right\}\right] \quad [\because X_K\text{'s are independent}]$$

$$= \left\{\phi\left(\frac{t}{\sqrt{n}}\right)\right\}^n \quad [\because X_K\text{'s are identical}]$$

$$= \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n \longrightarrow e^{-t^2/2} \text{ as } n\to\infty,$$

and $\phi(t)$ is the CF of $\frac{X_K - p}{\sqrt{p(1-p)}}$, &

since $e^{-t^2/2}$ is the CF of $N(0,1)$ RV, we have from continuity theorem,

$$P\left\{\frac{S_n-np}{\sqrt{npq}} \le x\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt = \Phi(x) \quad \forall\, x\in\mathbb{R}$$

$$\therefore \frac{S_n-np}{\sqrt{npq}} \xrightarrow{L} X \sim N(0,1)$$

$$\therefore S_n = \sum_{K=1}^{n} X_K \xrightarrow{L} \sqrt{npq}\,(X) + np \sim N(np, npq).$$

Alternative Proof:— Since X_K's are Bernoulli R.V.s $\therefore X_K$'s are i.i.d. with

$E(X_K) = p$ and $Var(X_K) = pq \le \frac{1}{4} < \infty$.

By Lindeberg – Levy theorem (CLT)

$$\frac{\sqrt{n}\,(\overline{X}_n - p)}{\sqrt{pq}} \xrightarrow{L} X \sim N(0,1)$$

i.e. $\frac{S_n - np}{\sqrt{npq}} \sim AN(0,1)$.

Scanned by CamScanner

<u>Alternative proof</u> :– PMF of S_n is given by,

$$p(x) = \frac{n!}{x!\,(n-x)!}\, p^x q^{n-x}$$

$$\simeq \frac{\sqrt{2\pi}\cdot e^{-n}\, n^{n+\frac{1}{2}}\, p^x q^{n-x}}{\sqrt{2\pi}\, e^{-x}\, x^{x+\frac{1}{2}}\, \sqrt{2\pi}\, e^{-(n-x)}\, (n-x)^{(n-x)+\frac{1}{2}}}$$

$$= \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}}$$ [By Stirling's approximation]

Define, $\delta = \dfrac{x-np}{\sqrt{npq}}$

$$\Rightarrow x = np + \delta\sqrt{npq}$$

$$\Rightarrow n-x = nq - \delta\sqrt{npq}$$

$$\therefore \frac{x}{np} = 1 + \delta\sqrt{\frac{q}{np}}$$

$$\therefore \frac{n-x}{nq} = 1 - \delta\sqrt{\frac{p}{nq}}$$

$$\therefore \binom{n}{x} p^x q^{n-x} \simeq \frac{1}{\sqrt{2\pi npq}} \left(1+\delta\sqrt{\frac{q}{np}}\right)^{-x-\frac{1}{2}}, \left(1-\delta\sqrt{\frac{p}{nq}}\right)^{-(n-x)-\frac{1}{2}}$$

$$\therefore \ln p(x) = c - \left(x+\frac{1}{2}\right)\ln\left(1+\delta\sqrt{\frac{q}{np}}\right) - \left(n-x+\frac{1}{2}\right)\ln\left(1-\delta\sqrt{\frac{p}{nq}}\right)$$

where $c = \ln \dfrac{1}{\sqrt{2\pi npq}}$

$$\Rightarrow \ln p(x) = c - \left(np + \delta\sqrt{npq} + \frac{1}{2}\right)\left\{\delta\sqrt{\frac{q}{np}} - \delta^2\frac{q}{2np} + \delta^3\frac{q^{3/2}}{3(np)^{3/2}} - \cdots\right\}$$
$$+ \left(nq - \delta\sqrt{npq} + \frac{1}{2}\right)\left\{\delta\sqrt{\frac{p}{nq}} + \delta^2\frac{p}{2nq} + \delta^3\frac{p^{3/2}}{3(nq)^{3/2}} + \cdots\right\}$$

$$= c + \left\{-\delta\sqrt{npq} + \delta\sqrt{npq}\right\} + \left\{-(\delta^2 q + \delta^2 p) + \left(\delta^2\frac{q}{2} + \delta^2\frac{p}{2}\right)\right.$$
$$+ \left\{\left(-\frac{\delta}{2}\sqrt{\frac{q}{np}} + \frac{\delta}{2}\sqrt{\frac{p}{nq}}\right) + \left(\delta^3\frac{q^{3/2}}{2\sqrt{np}} - \delta^3\frac{p^{3/2}}{2\sqrt{nq}}\right)\right.$$
$$\left.+ \left(-\delta^3\frac{q^{3/2}}{3\sqrt{np}} + \delta^3\frac{p^{3/2}}{3\sqrt{nq}}\right)\right\} +$$
$$\left\{\left(\frac{\delta^2 q}{4np} + \frac{\delta^2 p}{4nq}\right) - \left(\frac{\delta^4 p^2}{3np} + \frac{\delta^2 p^2}{3nq}\right) + \left(\frac{\delta^4 q^2}{4np} + \frac{\delta^4 p^4}{4nq}\right)\right\} + \cdots$$

assuming $\dfrac{|\delta|}{n^{1/6}} \to 0$ for large n,

$$\Rightarrow \frac{\delta}{\sqrt{n}} \to 0,\ \frac{\delta^3}{\sqrt{n}} \to 0,\ \frac{\delta^2}{n} \to 0,\ \frac{\delta^4}{n^{2/3}} \to 0 \Rightarrow \frac{\delta^4}{n} \to 0$$

$$\therefore \ln p(x) \simeq c - \frac{\delta^2}{2}$$

$$\Rightarrow p(x) = e^{c-\delta^2/2} = \frac{1}{\sqrt{2\pi npq}}\, e^{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2}.$$

Note: → 1. If for a sequence of random variables $\{X_n\}$ $\phi_{\bar{X}_n}(t) \to e^{it\mu}$, then we say WLLN holds for $\{X_n\}$ since $e^{it\mu}$ is the CF of a degenerate random variable X ∋ $P[X=\mu]=1$.

CLT as a generalisation of LLN: —

If $\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} X_k = \frac{S_n}{n}$ and

$\sigma^2 = V(X_k)$, $E(X_k)=\mu$.

$$\frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = S_n^*$$

is a standardised variate with mean zero and variance unity. If X_k's are i.i.d.

$$\therefore S_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

if $P[S_n^* \le x] \longrightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy$

then, $P[|S_n^*| \le x] \longrightarrow \Phi(x) - \Phi(-x)$

i.e. $P\left[|\bar{X}_n - \mu| \le \frac{x\sigma}{\sqrt{n}}\right] \longrightarrow \Phi(x) - \Phi(-x)$

∴ taking $\epsilon = \frac{x\sigma}{\sqrt{n}}$ ∴ $x = \frac{\epsilon\sqrt{n}}{\sigma}$.

∴ i.e. $P[|\bar{X}_n - \mu| \le \epsilon] \to \Phi\left(\frac{\epsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\epsilon\sqrt{n}}{\sigma}\right)$

$\to 1$ as $n \to \infty$

2. CLT may be looked as a generalisation of LLN since CLT gives the exact probability for large n, that $\{a \le \bar{X}_n \le b\}$ but LLN gives the limiting value of this probability.

Lyapunov's Form: →

$\{X_k\}$ is a sequence of independent random variables. WLG, we assume $E(X_k)=0$ and $V(X_k) = \sigma_k^2 < \infty$ then $V(S_n) = V\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \sigma_k^2 = \sigma_n^2$, and $\sigma_n^2 < \infty$.

Lyapunov Limit theorem : →

(Only Statement) Let $\{X_n\}$ be a sequence of independent random variables with $E(X_n) = \mu_n$, $V(X_n) = \sigma_n^2$,

$$E|X_n - \mu_n|^{2+\delta} < \infty \text{ for } \delta > 0$$

Define, $S_n = \sum_{k=1}^{n} X_k$,

Then

$$P\left[\frac{S_n - \sum_{k=1}^{n}\mu_k}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}} \le x\right] \to \Phi(x)$$

(distribution function of normal $(0,1)$)

whenever $n \to \infty$

Provided,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} E|X_k - \mu_k|^{2+\delta}}{\left(\sum_{k=1}^{n}\sigma_k^2\right)^{\frac{2+\delta}{2}}} = 0$$

known as Lyapunov's condition.

The cases $\delta > 1$ can be reduced to the case $\delta = 1$, Thus it is enough to consider $0 < \delta < 1$.

A specific condition to Lyapunov's condition: —

$$\frac{\left[\sum_{k=1}^{n} \rho_k^3\right]^{1/3}}{\left[\sum_{k=1}^{n}\sigma_k^2\right]^{1/2}} \to 0 \quad \text{as } n \uparrow \infty \qquad \rho_k^3 = E|(X_k - \mu_k)|^3.$$

Verify De-Moivre Laplace Theorem using Lyapunov's Theorem : —

Let $\{X_n\}$ be a sequence of independent Bernoulli random variable with,

$$P\{X_n = 1\} = p = 1 - P\{X_n = 0\}$$

Define, $S_n = \sum_{k=1}^{n} X_k$,

$$E(X_k) = 1 \times p + 0 \times (1-p) = p$$

$$E(X_k^2) = 1 \times p + 0 \times (1-p) = p$$

$$\sigma_k^2 = p - p^2 = pq.$$

$$\therefore E|X_k - \mu_k|^3 = |1-p|^3 p + |0-p|^3(1-p) = q^3 p + p^3 q$$

$$\sum_{k=1}^{n} E|X_k - \mu_k|^3 = n(q^3p + p^3q)$$

$$\therefore \left(\sum_{k=1}^{n}\sigma_k^2\right)^{3/2} = (npq)^{3/2}$$

$$\therefore \frac{\sum_{k=1}^{n} E|X_k - \mu_k|^3}{\left(\sum_{k=1}^{n}\sigma_k^2\right)^{3/2}} = \frac{n(p^3q + q^3p)}{(npq)^{3/2}} = \text{Constant}\cdot\frac{1}{\sqrt{n}} \to 0 \text{ as } n \to \infty$$