SAMPLING DISTRIBUTIONS

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

RANDOM SAMPLING AND SAMPLING DISTRIBUTION

Definition of Some Terms : —

1. Parameter : ~ A constant which changes its value from one situation to another. Specially, it is denoted by θ. A parameter labels a distribution uniquely.

2. Parameter Space : ~ Set of all admissible values of the parameter, denoted by Θ

Example : — i) $X \sim N(\mu, 1)$
μ = Parameter,
$\mathbb{R}$ = Parameter space.

ii) $X \sim N(\mu, \sigma^2)$
(μ, σ) = Parameter (vector valued)
Parameter Space = $\{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$

iii) $X \sim Bin(n, p)$
(n, p) = Parameter
Parameter space = $\{(n, p) : n \in \mathbb{N}, 0 < p < 1\}$

3. Labelling Parameter : ~ Suppose X is normally distributed with mean μ and s.d. unity. Then the parameter μ labels the distribution uniquely and hence termed as labelling parameter.

On the other hand, the parameter $\xi_{1/2}$, the median of a distribution though reflects a feature (regarding location) of the distribution, but it fails to label the distribution. But in case of one parameter Cauchy distribution with median θ, which labels the distribution.

Thus if a random variable X has distribution Function F, where the distribution is labelled or, indexed by the parameter θ. We denote the distn. by $F_\theta(\cdot)$.

Family of Distribution : —

Let X be a random variable having distribution function F_θ, $\theta \in \Theta$, then $\{F_\theta(\cdot) : \theta \in \Theta\}$ is said to be a family of distribution function, similarly, one may define a family of PDF or PMF's namely $\{f_\theta : \theta \in \Theta\}$, where $f_\theta(\cdot)$ is the PDF or PMF of X.

Example : — $\{\Phi(x-\mu) : \mu \in \mathbb{R}\}$
is a family of normal distribution with mean μ and s.d. unity.

4. Random Sample: — If $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variable each having distribution function F then $(X_1, X_2, \ldots, X_n)$ constitutes a random sample drawn from F.

5. Sample Space: — Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from a distribution having distribution function F. Suppose $(x_1, x_2, \ldots, x_n)$ is the realization on $(X_1, X_2, \ldots, X_n)$ then $(x_1, x_2, \ldots, x_n)$ is said to be a sample point. Clearly, these sample points may vary from one sampling location to another. The totality of all such sample points constitutes the sample space, commonly denoted by $\mathcal{X}$.

Example: Suppose, we have a random sample of size 2 from $N(\mu, 1)$ distribution. Then the sample space will be $\mathbb{R}^2$.

6. Statistic: — Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from a population having distribution function $F_\theta(\cdot)$. suppose $T(X_1, X_2, \ldots, X_n)$ is a measurable function $\ni$ $T: \mathbb{R}^n \to \mathbb{R}^k$ if $k=1$, T is said to be a real valued statistic and for $k>1$, T will be a vector valued statistic.

In simple words, statistic is a function of sample observation which is independent of any unknown parameters, i.e. here T does not depend on the labelling parameter θ.

Example: Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from $N(\mu, 1)$ population.

Here, the sample mean $\bar{X}$ is a statistic, we know, $\bar{X} \sim N(\mu, \frac{1}{n}) \Rightarrow \sqrt{n}(\bar{X}-\mu) \sim N(0,1)$.

It is to be noted that unless μ is specified, $\sqrt{n}(\bar{X}-\mu)$ would not be a statistic.

Once μ is specified as 2, $\sqrt{n}(\bar{X}-2)$ becomes a statistic.

Some real valued statistic are sample mean $\bar{X}$, sample Range R, sample s.d. 's'.

$X_{(1)}$, the minimum of the sample observation,

$X_{(n)}$, the maximum of the sample observation,

where as $(\bar{X}, s)$, $(X_{(1)}, X_{(n)})$, $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ are vector valued statistic.

[7]. Sampling Distribution : — The probability distribution of any statistic is termed as sampling distribution.

(2006) In a problem of parametric inference the population feature of interest can objectively be written as a function of labelling parameter, say $\gamma(\theta)$, where, an observation $X_1 \sim F_\theta$, which is not known completely except the form of F.

Now in order to guess $\gamma(\theta)$ [Problems of estimation] or, to validate any conjuctured regarding $\gamma(\theta)$ [Problem of hypothesis testing]. We proceed with a specific statistic and make use of its sampling distribution. In such inferential problem we always associate a measure of error with the conclusion where this error is nothing but the sampling error. As a measure the s.d. of the sampling distribution of the statistic would serve the purpose and this is termed as the standard error.

▨ Exhibiting a sampling distribution in case of sampling from a finite identifiable population : —

Here the term identifiability means that the population units can easily be distinguished.

[8]. Simple Random Sampling : — Suppose we have a finite identifiable population of size N $(u_1, u_2, \ldots, u_N)$, where u_α is the αth member of the population. By a sample we mean, a non-empty collection of units from $(u_1, u_2, \ldots, u_N)$ with, or, without repetations. Here the sampling procedure may be subjective (purposive sampling, deliberate sampling, haphazard sampling) or, objective (Probabilistic, non-probabilistic, mixed).

The probabilistic sampling may be an equal probability sampling where each of the possible sample has the same probability to occur (or, every unit of the population has the same probability to be included in the sample) or, an unequal probability sampling.

• Definition : → Simple random sampling (SRS) is an equal probability sampling. An SRS may be drawn with replacement termed as SRSWR or, without replacement termed as SRSWOR.

SRSWR :— Suppose a sample of n units is drawn from the population of size N one by one with replacement. Clearly, number of possible samples is N^n on each has probability $\frac{1}{N^n}$ to occur.

SRSWOR :— Suppose a sample of n units is drawn at random one by one without replacement from the population of size N. If we ignore the order of the units in the sample. The number of possible sample will be $\binom{N}{n}$ and each has probability $\frac{1}{\binom{N}{n}}$ to occur.

On the other hand, if the order is taken into account the number of possible sample is $(N)_n$ and each has the probability $\frac{1}{(N)_n}$ to occur.

Suppose an SRSWOR of size 3 is drawn from a population of size 5.

$(u_1, u_2, u_3, u_4, u_5)$

Let us ignore the order of the units in the sample.

Further assume that the variate values of the population units are 6, 8, 4, 6, 8, respectively.

Let s be a typical sample and $\bar{x}(s)$ be the sample mean. Then we have the following sampling distribution of the sample mean.

Serial No.	s	$\bar{x}(s)$
1	(u_1, u_2, u_3)	6
2	(u_1, u_2, u_4)	6.67
3	(u_1, u_2, u_5)	7.33
4	(u_1, u_3, u_4)	5.33
5	(u_1, u_3, u_5)	6
6	(u_1, u_4, u_5)	6.67
7	(u_2, u_3, u_4)	6
8	(u_2, u_3, u_5)	6.67
9	(u_2, u_4, u_5)	7.33
10	(u_3, u_4, u_5)	6

$\bar{x}$	
5.33	1/10
6	2/5
6.67	3/10
7.33	1/5

Distribution of a Statistic :

- Simple Random Sampling :— From a population of N units, select [2004] one by one giving equal probability to all units. One makes a note of the unit selected and return it to the population. If this operation is performed n times we get a simple random sample of n units, selected with replacement. Not returning the unit on the units selected and selecting a further unit with equal chance from the units that remain in the population, then we get a simple random sample selected without replacement.

Definition :→ If each unit of the population has an equal probability of being selected at each drawing, then the sampling is called simple random sampling.

- Random Sampling from a probability distribution :— We have [2004] defined a random sampling in the context of a finite popln. but there will be many cases where the popln. has to be considered as infinite on hypothetical. For an infinite popln., sampling with replacement and without replacement are practically equivalent. Any two or more of the sampling values and independently distributed under SRSWR and consequently for large N or for hypothetical popln., approximately independently distributed under SRSWOR. We characterise an infinite population in terms of probability distn. of some R.V. X. By random sampling of size n, from the probability distn. we mean a set of n random variables $X_1, X_2, \ldots, X_n$ such that $X_1, \ldots, X_n$ are independently distributed and each have the same probability distn. of X.

Remark :→ Sampling from a probability distn. is sometimes refer to as sampling from an infinite population. Since, one can obtain samples of any size one desires even if the study variable takes only finite no. of values.

- Parameter :— Our primary interest will be in knowing the [C.U] value of different measures of characteristic of the population distribution of X like its mean, standard deviation. The measure of this type calculated on the basis of popln values of X is called a parameter. In a narrow sense, a parameter is a constant that occurs in the probability distn. of the variable. As for example, λ is a parameter of Poisson variable X, $X \sim P(\lambda)$, and (μ, σ) are parameters of the normal variable X, $X \sim N(\mu, \sigma^2)$.

[CU]

- **Statistic :** — Statistic is a function of sample values which is itself an observable (or, measurable) random variable which does not contain any parameter.

If $x_1, x_2, \ldots, x_n$ be a random sample from a popln. with mean μ and variance σ^2. Then $\bar{x}-\mu$ and $\frac{s^2}{\sigma^2}$ are not statistic, since they contain parameter and are not observable or measurable. But if $\mu=\mu_0$ (known) then $(\bar{x}-\mu_0)$ is a statistic. Clearly, the sample mean $\bar{x}$, the sample variance $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ and $x_{(1)}, x_{(n)}$ are examples of statistic.

[CU]

- **Sampling Distribution and Standered Error of a Statistic :-**

Let $x_1, x_2, \ldots, x_n$ be a random sample from a population and let $T=T(x_1, \ldots, x_n)$ be a statistic. Since, the set of popln. members included in different samples may be different, the value of the statistic T is liable to vary from one sample to another. These differences in the values of the Statistic T are called Sampling fluctuation of the statistic.

If a no. of samples each of size n are drawn from the popln. and if for each sample the value of the statistic T is calculated, then a collection of values of the statistic will be obtained. If the no. of samples, N is large, this may be arranged into a frequency table.

Values of T	Frequency	Relative Freq.	Probability
t_1	f_1	f_1/N	$P[T=t_1]=f_1/N$
t_2	f_2	f_2/N	$P[T=t_2]=f_2/N$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_k	f_k	f_k/N	$P[T=t_k]=f_k/N$.

Note that $\lim_{N\to\infty}\frac{f_i}{N}=P[T=t_i]$, by Statistical defn. of Probability.

The approximate probability distn. or the frequency distn. of statistic that would be obtain if the no. of samples N is indefinitely large, is called the sampling distn. of the statistic.

If $(x_1, x_2, \ldots, x_n)$ is a random sample from popln. with known D.F. $F(x)$, then the probability distn. of T can be obtained by transformation technique or, any other technique.

Ex: If $x_1, x_2, \ldots, x_n$ is a r.s. from $N(\mu, \sigma^2)$, then $x_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i=1(1)n$ and it can be shown that $\bar{x}\sim N(\mu, \frac{\sigma^2}{n})$. Hence, the sampling distribution of $T=\bar{x}$ is $N(\mu, \frac{\sigma^2}{n})$.

[CU]

- **Standard Error :** — Sampling distribution of a statistic(T) is a probability distribution and it has mean, variance and higher order moment like, any other probability distn. The s.d. of the sampling distn of the statistic T is called the standard error of the ~~statistic~~ statistic.

Mathematically, standard Error of $T = \sqrt{Var(T)}$

$$= \left[E\{T - E(T)\}^2\right]^{1/2}$$

Standard error gives the variability of the values of the statistic T about its mean the average. $E(T) = \theta$, say. S.E. (T) gives an idea of the error that one could commit, using the values of the statistic (T), in estimating the parameter $\theta = E(T)$. The smaller the S.E., the better the prediction or guess.

- Remark:— Let $X_1, X_2, \ldots, X_n$ be a random sampling from a popln. with mean μ and variance σ^2. Here, the popln. variability is measured by $\sigma^2 = var(X_i)$, $\forall\, i$. Hence, the standard deviation of the population is σ. Note that, $X_1, \ldots, X_n$ are i.i.d. R.V.'s with mean μ and variance σ^2. Consider the statistic T. Then Var (T) is, in general, a function of population standard deviation (σ), i.e. the $S.E.(T) = \sqrt{V(T)}$ is, in general, a function of popln standard deviation(σ).

In particular, if $T = \bar{X}$,

then $Var(T) = Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{V(X_1)}{n}$.

$$S.E\,(T) = \sqrt{V(T)} = \frac{\sigma}{\sqrt{n}} = \frac{\{\text{s.d. of the popln.}\}}{\sqrt{n}}.$$

Clearly, the S.D. (σ) of the popln gives a measure of popln. variability where as S.E(T) gives a measure of variability of the sampling distribution of the statistic T.

× ——— ×

Some Basic Distributions useful in finding Sampling Distribution of a statistic :

A. χ^2 - distribution [Chi-square Distribution] :

- **Definition:** A random variable χ^2 is said to have χ^2-distribution with n degrees of freedom if its PDF is given by

$$f_{\chi^2}(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \cdot e^{-\frac{x}{2}} x^{\frac{n}{2}-1} & , \ 0 < x < \infty \\ 0 & , \ \text{OW} \end{cases}$$

where $n > 0$.

Notation: $\chi^2 \sim \chi^2_n$

$\Rightarrow \chi^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{n}{2}\right)$.

- **Derivation of χ^2-distribution:** Let $(X_1, X_2, \ldots, X_n)$ be an i.i.d. random sample drawn from $N(0,1)$ population then the statistic $\sum_{i=1}^{n} X_i^2$ follows χ^2-distribution with n degree of freedom. We denote, $\sum_{i=1}^{n} X_i^2 \sim \chi^2_n$.

Proof: The joint PDF of $(X_1, X_2, \ldots, X_n)$ is given by,

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} \quad , \ x_i \in \mathbb{R}$$

$; (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

Consider the Polar transformation

$$x_1 = r\cos\theta_1$$
$$x_2 = r\sin\theta_1 \cos\theta_2$$
$$x_3 = r\sin\theta_1 \sin\theta_2 \cos\theta_3$$
$$\vdots$$
$$x_{n-1} = r\sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-2} \cos\theta_{n-1}.$$
$$x_n = r\sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-2} \sin\theta_{n-1}.$$

where, $0 < r < \infty$, $0 < \theta_i < \pi$, $i = 1(1)\overline{n-2}$, $0 < \theta_{n-1} < 2\pi$

Here, $\sum_{i=1}^{n} x_i^2 = r^2$

and transformation is one-to-one;

The jacobian J is given by,

$$|J| = r^{n-1} \cdot (\sin\theta_1)^{n-2} \cdot (\sin\theta_2)^{n-3} \ldots \ldots \sin\theta_{n-2}.$$

The PDF of $(r, \theta_1, \theta_2, \dots, \theta_{n-1})$ is

$$f_{r\theta_1\theta_2\cdots\theta_{n-1}}(r,\theta_1,\theta_2\cdots\theta_{n-1}) = \begin{cases} \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-r^2/2} \cdot r^{n-1} \cdot (\sin\theta_1)^{n-2} \cdot (\sin\theta_2)^{n-3} \cdots\cdots (\sin\theta_{n-2}) ; & \text{if } 0<r<\infty,\ 0<\theta_i<\pi,\ i=1(1)\overline{n-2},\ 0<\theta_{n-1}<2\pi. \\ 0 & ; \text{elsewhere} \end{cases}$$

This shows that the R.V.'s $r, \theta_1, \theta_2, \dots, \theta_{n-1}$ are independently distributed.

The PDF of r is, $f_r = \begin{cases} c.e^{-\frac{1}{2}r^2} r^{n-1} & , 0<r<\infty \\ 0 & , \text{OW} \end{cases}$

Now, $\chi^2 = \sum_{i=1}^{n} x_i^2 = r^2$,

$P[r^2 \le x] = P[r \le \sqrt{x}] = F_r(\sqrt{x})$

Hence the PDF of χ^2 is $f_{\chi^2}(x) = \begin{cases} f_r(\sqrt{x}) \cdot \left|\frac{dr}{dx}\right| & , 0<r<\infty \\ 0 & , \text{OW} \end{cases}$

$\left[\frac{dr}{dx} = \frac{d}{dx}(\sqrt{x}) = \frac{1}{2\sqrt{x}}\right]$

$$= \begin{cases} c.e^{-x/2} \cdot x^{\frac{n-1}{2}} \cdot \frac{1}{2\sqrt{x}} & , \text{if } 0<x<\infty \\ 0 & , \text{OW} \end{cases}$$

$$= \begin{cases} c_1 . e^{-x/2} . x^{\frac{n}{2}-1} & , 0<x<\infty \\ 0 & , \text{OW} \end{cases}$$

As, $1 = \int_0^\infty f_{\chi^2}(x)\,dx$

$= c_1 \int_0^\infty e^{-x/2} . x^{n/2-1}\,dx$

$= c_1 . \Gamma\left(\frac{n}{2}\right) . 2^{n/2}$

$\Rightarrow c_1 = \frac{1}{2^{n/2} . \Gamma\left(\frac{n}{2}\right)}$

$\therefore$ The PDF of $\chi^2 = \sum_{i=1}^{n} x_i^2$ is

$$f_{\chi^2_n}(x) = \begin{cases} \frac{e^{-x/2} . x^{n/2-1}}{2^{n/2} . \Gamma\left(\frac{n}{2}\right)} & , \text{if } 0<x<\infty \\ 0 & , \text{OW} \end{cases}$$

Remark:—

Degree of freedom in any expression is the no. of unrestricted variables in the expression.

In $\sum_{i=1}^{n}(x_i-\bar{x})^2$, we have n variables $x_i-\bar{x}$, $i=1(1)n$, not all independent but restricted by the relation, $\sum_i (x_i-\bar{x})=0$.

Therefore, $\sum_{i=1}^{n}(x_i-\bar{x})^2$ has $(n-1)$ degrees of freedom.

Consider the expression $\sum_{i=1}^{n} x_i^2$ subject to the restriction $\sum_{i=1}^{n} x_i=0$ and $x_1+x_n=0$. clearly, the expression has two restriction and hence, $(n-2)$ quantities are unrestricted or independent. Therefore the expression has $(n-2)$ degree of freedom.

In general, as a rule, the degrees of freedom

= (the no. of quantities involves in the expression) − (the no. of linear restriction given).

- Alternative method:—

Proof:→ The joint PDF of $(X_1, X_2, \dots, X_n)$ given by,

$$f_{X_1X_2\dots X_n}(x_1,x_2,\dots,x_n)=\left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2}, \quad (x_1,x_2,\dots,x_n)\in\mathbb{R}^n$$

Consider the following transformation,

$$(x_1,x_2,\dots,x_n)\to(r,\theta_1,\theta_2,\dots,\theta_{n-1})$$

$x_1 = r\cos\theta_1$,

$x_2 = r\sin\theta_1\cos\theta_2$,

$x_3 = r\sin\theta_1\sin\theta_2\cos\theta_3$,

⋮

$x_{n-1} = r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\cos\theta_{n-1}$.

$x_n = r\sin\theta_1\sin\theta_2\cdots\cdots\sin\theta_{n-2}\sin\theta_{n-1}$.

$\therefore \sum_{i=1}^{n}x_i^2=r^2$

$0<r<\infty$

$0<\theta_i<\pi \quad \forall\ i=1(1)\overline{n-2}$.

$0<\theta_{n-1}<2\pi$.

$$J\left(\frac{x_1,x_2,\dots,x_n}{r,\theta_1,\theta_2,\dots,\theta_{n-1}}\right)=r^{n-1}\sin^{n-2}\theta_1\sin^{n-3}\theta_2\cdots\sin\theta_{n-2}$$

Note, $\beta(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}\,dx$

$$= 2\int_0^{\pi/2} \sin^{2m-1}\theta \cos^{2n-1}\theta\,d\theta \qquad \left|\begin{array}{l} x = \sin^2\theta \\ dx = 2\sin\theta\cos\theta\,d\theta \end{array}\right.$$

$$\therefore \beta\left(\frac{m+1}{2}, \frac{n+1}{2}\right) = 2\int_0^{\pi/2} \sin^m\theta \cos^n\theta\,d\theta$$

Note, $\int_0^{\pi} \sin^k\theta\,d\theta$

$$= \int_0^{\pi} \cos^k\left(\frac{\pi}{2}-\theta\right)d\theta$$

$$= \int_{-\pi/2}^{\pi/2} \cos^k\phi\,d\phi \qquad \left|\begin{array}{l} \frac{\pi}{2}-\theta = \phi \\ -d\theta = d\phi \end{array}\right.$$

$$= 2\int_0^{\pi/2} \cos^k\phi\,d\phi$$

$$= \beta\left(\frac{1}{2}, \frac{k+1}{2}\right)$$

Joint PDF of $r, \theta_1, \theta_2, \ldots, \theta_{n-1}$ is

$f_{r,\theta_1,\theta_2,\ldots,\theta_{n-1}}(r, \theta_1, \theta_2, \ldots, \theta_{n-1})$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{r^2}{2}}\, r^{n-1} \sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \ldots\ldots \sin\theta_{n-2}$$

$0 < r < \infty$
$0 < \theta_i < \pi$
$\forall\ i = 1(1)\,\overline{n-2}$
$0 < \theta_{n-1} < 2\pi$

$f_\chi(r)$ = PDF of r

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-r^2/2} \cdot r^{n-1} \cdot 2\pi \cdot \int_0^{\pi} \sin^{n-2}\theta_1\,d\theta_1 \int_0^{\pi} \sin^{n-3}\theta_2\,d\theta_2 \cdots$$

$$\cdots \int_0^{\pi} \sin\theta_{n-2}\,d\theta_{n-2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-r^2/2} \cdot 2\pi \cdot r^{n-1}\, \beta\left(\frac{1}{2}, \frac{n-1}{2}\right)\beta\left(\frac{1}{2}, \frac{n-2}{2}\right) \cdots\cdots \beta\left(\frac{1}{2}, 1\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-r^2/2} \cdot 2\pi \cdot \left(\Gamma\tfrac{1}{2}\right)^{n-2} \cdot \frac{1}{\Gamma\frac{n}{2}} \cdot r^{n-1}$$

$$= \frac{2 \cdot \pi \cdot \pi^{\frac{n-2}{2}}}{2^{n/2} \cdot \pi^{n/2}} \cdot \frac{1}{\Gamma\frac{n}{2}} \cdot e^{-r^2/2} \cdot r^{n-1}$$

$$= \frac{1}{2^{\frac{n-2}{2}} \cdot \Gamma\frac{n}{2}} \cdot e^{-r^2/2} \cdot r^{n-1} \cdot I_r(0,\infty)$$

$$\therefore f_{\chi}(x) = \frac{1}{2^{\frac{n-2}{2}} \cdot \Gamma\frac{n}{2}} \cdot e^{-x^2/2} \cdot x^{n-1} \, I_x(0,\infty)$$

Let, $z = \chi^2$,

$\sqrt{z} = \chi$, $0 < z < \infty$

$$\therefore |J| = \left|\frac{dx}{dz}\right| = \frac{1}{2\sqrt{z}}$$

$$\therefore f_{\chi^2}(z) = \frac{1}{2^{\frac{n}{2}-1}\Gamma\frac{n}{2}} e^{-\frac{1}{2}z} \cdot z^{\frac{n-1}{2}} \cdot \frac{1}{2\sqrt{z}}$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot e^{-\frac{1}{2}z} z^{\frac{n}{2}-1} \, I_z(0,\infty)$$

Note that, for $\frac{1}{2} = \alpha$, $\frac{n}{2} = p$, χ_n^2 is nothing but a Gamma distribution. It is a specific case of 2 parameter gamma distribution. Thus χ_2^2 is an exponential distribution with mean 2.

Features:- Since it is a specific Gamma distribution, Chi-square distribution is positively skewed and leptokurtic and it has a reproductive property given by,

$$\sum_{i=1}^{m} x_i \sim \chi^2_{\sum_{i=1}^{m} n_i}$$

Calculation of mean & variance:-

$$E(\chi^2) = \int_0^\infty \frac{z \cdot e^{-\frac{1}{2}z} z^{\frac{n}{2}-1}}{2^{n/2}\Gamma\frac{n}{2}} dz$$

$$= \frac{\Gamma\left(\frac{n}{2}+1\right) \cdot 2^{\frac{n}{2}+1}}{\Gamma\frac{n}{2} \cdot 2^{n/2}}$$

$$= n.$$

$$V(\chi^2) = \int_0^\infty \frac{z^2 \cdot e^{-\frac{1}{2}z} z^{n/2-1}}{\Gamma\frac{n}{2} \cdot 2^{n/2}} dz - E^2(\chi^2)$$

$$= \frac{\Gamma\left(\frac{n}{2}+2\right) \cdot 2^{\frac{n}{2}+2}}{\Gamma\frac{n}{2} \cdot 2^{n/2}} - n^2$$

$$= \frac{\left(\frac{n}{2}+1\right) n \cdot \frac{1}{2} \cdot \cancel{2^{n/2}} \cdot 2^2 \cdot \cancel{\Gamma\frac{n}{2}}}{\cancel{\Gamma\frac{n}{2}} \cdot \cancel{2^{n/2}}} - n^2$$

$$= (n+2)n - n^2$$

$$= 2n.$$

MGF:— $M_{\chi^2}(t) = (1-2t)^{-n}$ if $t < \frac{1}{2}$.

Moments:— $\mu_r' = E(\chi^2)^r$

$$= \frac{\int_0^\infty e^{-x/2} x^{n/2+r-1}\,dx}{2^{n/2}\Gamma(\frac{n}{2})}$$

$$= \frac{2^{n/2+r}\,\Gamma(\frac{n}{2}+r)}{2^{n/2}\,\Gamma(\frac{n}{2})}, \text{ if } r > -\frac{n}{2}$$

$$= \frac{2^r \cdot \Gamma(n/2+r)}{\Gamma(n/2)}, \text{ if } r > -\frac{n}{2}.$$

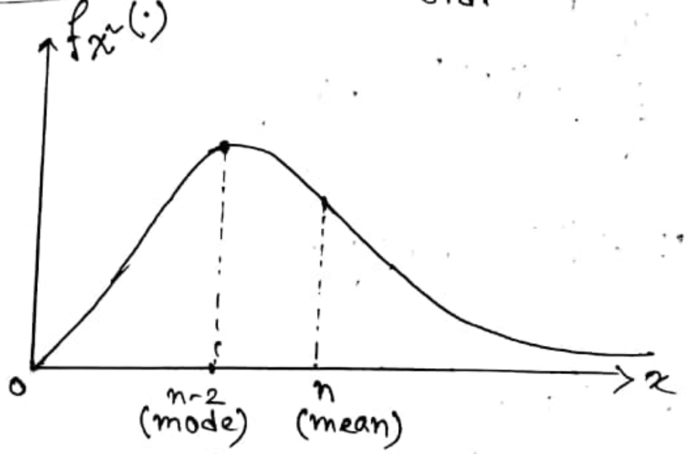
For, $r=1$, $\mu_1' = \frac{2\Gamma(n/2+1)}{\Gamma(n/2)} = n = E(\chi^2)$

and $\mu_2' = E(\chi^2)^2 = \mu_2' = n(n+2)$

$E[\chi^2_n] = n$ and $var[\chi^2_n] = 2n$.

Mode:— $Mode(\chi^2_n) = n-2$, if $n>2$.

Skewness:— $S_K = \frac{\text{Mean} - \text{mode}}{\text{s.d.}} = \frac{n-(n-2)}{\sqrt{2n}} = \sqrt{\frac{2}{n}} > 0$



Hence, χ^2-distribution is positively skewed and leptokurtic.

Note that, as $n \to \infty$, $S_K \to 0$, i.e. χ^2_n becomes a symmetric distribution.

☆ Ex.1. If $p = P[\chi_2^2 > x_0]$, then show that $x_0 = 2\ln\left(\frac{1}{p}\right)$.

Soln → $p = P[\chi_2^2 > x_0]$

$$= \int_{x_0}^{\infty} \frac{1}{2^{2/2}\Gamma\left(\frac{2}{2}\right)} e^{-\frac{x}{2}} \cdot x^{\frac{2}{2}-1} \, dx$$

$$= \int_{x_0}^{\infty} \frac{1}{2} \cdot e^{-x/2} dx$$

$$= e^{-x_0/2}$$

$$\Rightarrow -\frac{x_0}{2} = \ln(p)$$

$$\Rightarrow x_0 = 2\ln\left(\frac{1}{p}\right)$$

☆ Ex.2. If $p = P[T \le t]$, where T is an absolutely continuous R.V. Then show that $-2\log_e p \sim \chi_2^2$.
Also find the distribution of $-2\log_e p$, where
i) $p = P[T \geq t_0]$, ii) $p = P[|T| \geq |t|]$

Soln → $p = P[T \le t]$

$$= F_T(t) \sim U(0,1).$$

Then $f(p) = \begin{cases} 1, & 0 < p < 1 \\ 0, & \text{ow} \end{cases}$

The PDF of $Y = -2\log_e p$ is

$$f_Y(y) = \begin{cases} 1 \cdot \left| \frac{d}{dy}\left(e^{-y/2}\right) \right|, & 0 < y < \infty \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{-y/2}, & 0 < y < \infty \\ 0, & \text{ow} \end{cases}$$

$$\Rightarrow Y = -2\log_e p \sim \chi_2^2.$$

Note that $P[T \geq t] = 1 - F_T(t) \sim U(0,1)$.
and $P[|T| \geq |t|] = P[U \geq u]$, $U = |T|$,

$$= 1 - F_U(u) \sim U(0,1).$$

☆ Ex.3. If $p_x = \frac{1}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)} \int_0^x e^{-t/2} t^{n/2-1} dt$, S.T. $x < \frac{n}{1-p_x}$.
[C.U. 2004]

Soln → $p_x = P[\chi_n^2 \le x]$

$$= 1 - P[\chi_n^2 > x]$$

$$> 1 - \frac{E(\chi_n^2)}{x} \quad \text{[By Markov's inequality]}$$

$$\Rightarrow p_x > 1 - \frac{n}{x}$$

$$\Rightarrow \frac{n}{x} > 1 - p_x \Rightarrow x < \frac{n}{1-p_x}$$

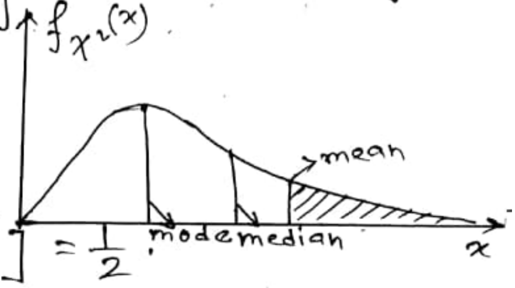
☆ **Ex. 4.** Show that $P[\chi_n^2 > \lambda] < \frac{n}{\lambda}$, $\lambda > 0$

Show that $P[\chi_n^2 > n] < \frac{1}{2}$.

soln. → For a non-negative R.V. χ_n^2, by Markov's inequality,

$$P[\chi_n^2 > \lambda] < \frac{E(\chi_n^2)}{\lambda}$$

$$\Rightarrow P[\chi_n^2 > \lambda] < \frac{n}{\lambda}.$$



Now, $P[\chi_n^2 > n] = P[\chi_n^2 > E(\chi_n^2)]$

$< P[\chi_n^2 > \text{med}(\chi_n^2)] = \frac{1}{2}$.

A Special Topic :—

- **Question** :— $X_1, X_2, \ldots, X_n$ are independent $N(0,1)$ variables. obtain the distribution of $\sum_{i=1}^{n} X_i^2$ subject to the following $m(<n)$ linearly independent constants

$$a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{in}X_n = 0 \quad , i = 1, 2, \ldots, m.$$

, where a_{ij}'s are constants, $j = 1(1)n$. Find distn. of $\sum_{i=1}^{n} X_i^2 / A\underset{\sim}{X} = \underset{\sim}{0}$

ANSWER :— Define $A = ((a_{ij}))_{m \times n}$

Given that the rows of A are linearly independent vectors. Hence, through row operations the rows of A can be reduced to mutually orthonormal vectors. Thus $\exists$ a non-singular matrix $m \times m$ $\ni$ $PA = B$ $\ni$ $B'B = I_m$

$$\therefore \; A\underset{\sim}{X} = \underset{\sim}{0}$$

$$\Leftrightarrow PA\underset{\sim}{X} = \underset{\sim}{0}$$

i.e. $B\underset{\sim}{X} = \underset{\sim}{0}$

Clearly, there exists a matrix $C_{\overline{n-m} \times n}$ $\ni$

$\begin{pmatrix} B \\ C \end{pmatrix}$ is an orthogonal matrix.

$\begin{pmatrix} B \\ C \end{pmatrix} = D$, say,

Consider the following orthogonal transformation, $\ni$

$$\underset{\sim}{Y} = D\underset{\sim}{X}$$

Here, $\underset{\sim}{x} \in \mathbb{R}^n$

$\Rightarrow \underset{\sim}{y} \in \mathbb{R}^n$.

$$\underset{\sim}{y}'\underset{\sim}{y} = \underset{\sim}{x}'D'D\underset{\sim}{x} = \underset{\sim}{x}'\underset{\sim}{x}$$

$$\left| J\left(\frac{x_1 \ldots x_n}{y_1 \ldots y_n}\right)\right| = 1.$$

Joint PDF of $\underset{\sim}{X}$

$$f_{\underset{\sim}{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2}$$

$Y_1, Y_2, \ldots, Y_n$ are independent $N(0,1)$ variable.

Partition $\underset{\sim}{Y}$ into

$$\underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}_{(1)} \\ \underset{\sim}{Y}_{(2)} \end{pmatrix}$$

Given $A\underset{\sim}{X} = \underset{\sim}{0} \Rightarrow B\underset{\sim}{X} = \underset{\sim}{0}$

Now, $\underset{\sim}{Y} = D\underset{\sim}{X} = \begin{pmatrix} B \\ C \end{pmatrix} \underset{\sim}{X} = \begin{pmatrix} B\underset{\sim}{X} \\ C\underset{\sim}{X} \end{pmatrix}$

$$\underset{\sim}{Y}_{(1)} = B\underset{\sim}{X} = \underset{\sim}{0}$$

Define, $\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{Y}'\underset{\sim}{Y} / A\underset{\sim}{X} = \underset{\sim}{0}$ is same as the distribution of $\underset{\sim}{Y}'\underset{\sim}{Y} / \underset{\sim}{Y}_{(1)} = \underset{\sim}{0}$

$$\Rightarrow \sum_{i=m+1}^{n} Y_i^2 \sim \chi^2_{n-m}.$$

Scanned by CamScanner

Some Important properties of Chi-square Distribution :—

(1) If $X \sim \chi^2_{n_1}$ and $Y \sim \chi^2_{n_2}$, independently, then—

(a) $X + Y \sim \chi^2_{n_1+n_2}$.

(b) $\frac{X}{X+Y} \sim$ First kind Beta $\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$

(c) $\frac{X}{Y} \sim$ Second kind Beta $\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$. [C.U. 2001]

(2) **CLT:** $\frac{\chi_n^2 - n}{\sqrt{2n}} \xrightarrow{L} Z \sim N(0,1)$, as $n \to \infty$

Derivation of the CLT.

[Hints:- Consider a sequence $\{X_n\}$ of iid R.V.'s each following χ^2- distribution with 1 degree of freedom. Then apply Lindeberg - Levy CLT.]

(3) $\sqrt{2\chi_n^2} \overset{a}{\sim} N(\sqrt{2n-1}, 1)$, as $n \to \infty$.

Proof:→ To show, $\sqrt{2\chi_n^2} - \sqrt{2n-1} \overset{a}{\sim} N(0,1)$

$$\Leftrightarrow \sqrt{2\chi_n^2} - \sqrt{2n-1} \xrightarrow{L} Z \sim N(0,1).$$

Note that,

$$P[\sqrt{2\chi_n^2} - \sqrt{2n-1} \le z]$$
$$= P[2\chi_n^2 \le (\sqrt{2n-1} + z)^2]$$
$$= P[2\chi_n^2 \le (2n-1) + 2\sqrt{2n-1}\cdot z + z^2]$$
$$= P[\chi_n^2 \le n - \tfrac{1}{2} + z\sqrt{2n-1} + \tfrac{z^2}{2}]$$
$$= P\left[\frac{\chi_n^2 - n}{\sqrt{2n}} \le z\sqrt{\frac{2n-1}{2n}} + \frac{z^2-1}{2\sqrt{2n}}\right]$$
$$\simeq P\left[\left(\frac{\chi_n^2 - n}{\sqrt{2n}}\right) \le z\right], \text{ for large } n.$$
$$\to P[Z \le z], \text{ as } n \to \infty = \Phi(z).$$

[As $n \to \infty$, by CLT, $\frac{\chi_n^2 - n}{\sqrt{2n}} \xrightarrow{L} Z \sim N(0,1)$]

Hence, $\sqrt{2\chi_n^2} - \sqrt{2n-1} \xrightarrow{L} Z \sim N(0,1)$.

[B]. t - distribution :—

- **Definition** :— Suppose X is a standard normal variate, and Y is a χ^2 variate having n-degree of freedom. Moreover, X and Y are independently distributed,

Define $t = \dfrac{X}{\sqrt{Y/n}}$

Then 't' is said to have a t-distribution with n-degrees of freedom if the PDF is of the form

$$f_t(t) = \frac{1}{\sqrt{n}\, \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}, \quad t \in \mathbb{R}.$$

We denote, $t \sim t_n$.

- **Derivation of the PDF** :— [C.U. 2000]

Joint PDF of X and Y are given by,

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} \cdot \dfrac{e^{-\frac{1}{2}(x^2+y)}}{2^{n/2} \cdot \Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1}, & x \in \mathbb{R},\ y \in \mathbb{R}^+ \\ 0, & \text{o/w} \end{cases}$$

Consider the transformation

$$(X, Y) \longrightarrow (t, Y), \text{ where } t = \frac{X}{\sqrt{Y/n}}$$

$$x \in \mathbb{R}, y \in \mathbb{R}^+ \Rightarrow t \in \mathbb{R}$$

$$x = t\sqrt{y/n}$$

Here, $|J| = J\left(\dfrac{x, y}{t, y}\right) = \sqrt{\dfrac{y}{n}}$.

Now,
$$f_{tY}(t,y) = \frac{e^{-\frac{1}{2}\left(\frac{t^2 y}{n} + y\right)}}{\sqrt{2\pi} \cdot 2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} \cdot |J|, \quad t \in \mathbb{R},\ y \in \mathbb{R}^+$$

$$= \frac{e^{-\frac{1}{2}y\left(1+\frac{t^2}{n}\right)}}{\sqrt{2\pi} \cdot 2^{n/2}\Gamma\left(\frac{n}{2}\right)} \cdot y^{\frac{n}{2}-1} \cdot \sqrt{\frac{y}{n}}$$

$$= \frac{1}{\sqrt{n} \cdot 2^{\frac{n+1}{2}} \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)} \cdot e^{-\frac{y}{2}\left(1+\frac{t^2}{n}\right)} \cdot y^{\frac{n+1}{2}-1}, \quad y \in \mathbb{R}^+,\ t \in \mathbb{R}$$

Hence, $f_t(t) = \displaystyle\int_0^\infty f_{tY}(t,y)\, dy$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n} \cdot 2^{\frac{n+1}{2}} \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\left\{\frac{1}{2}\left(1+\frac{t^2}{n}\right)\right\}^{\frac{n+1}{2}}}$$

$$= \frac{1}{\sqrt{n}\, \beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}, \quad t \in \mathbb{R}$$

Note :– The PDF of distribution t_1 is standard Cauchy.
t-distribution is symmetric about zero, bell-shaped distn..
Clearly, the median = mode = 0; and mean of the distribution if exists, depending on n, vanishes.

Moments :–

(a) If $t \sim t_n$, then $\mu_r' = E(t^r)$ exists, iff $r < n$.

(b) Note that, $f_t(-t) = f_t(t) \;\forall t$. [CU. 2004]

$\Rightarrow$ the distribution is symmetric about zero.

Hence for $n > 1$, $E(t) = 0$ and $\mu_{2r-1} = 0$, if they exists.

(c) $$\mu_{2r} = \mu_{2r}' = E(t^{2r}) = n^r E\left(\left(\frac{X^2}{Y}\right)^r\right)$$

$$= n^r E[(X^2)^r] E(Y^{-r})$$

$$= n^r \cdot \frac{2^r \cdot \Gamma\left(\frac{1}{2}+r\right)}{\Gamma\left(\frac{1}{2}\right)} \cdot \frac{2^{-r} \cdot \Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{n}{2}\right)} \text{ if } -\frac{1}{2} < r < \frac{n}{2} \Leftrightarrow -1 < 2r < n$$

$$= n^r \cdot \frac{\Gamma\left(\frac{1}{2}+r\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}, \quad -1 < 2r < n.$$

$E(t) = 0$, if $n > 1$

$\text{var}(t) = \mu_2 = \frac{n}{n-2}$, if $n > 2$.

☆ **Ex. 1.** If $X \sim t_n$, show that $Y = \frac{1}{1+\frac{X^2}{n}}$ has a beta distribution.

Soln. $\to$ PDF of X is :

$$f_X(x) = \frac{\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}}{\sqrt{n}\,\beta\left(\frac{n}{2}, \frac{1}{2}\right)}, \quad x \in \mathbb{R}$$

Here, $y = \frac{1}{1+\frac{x^2}{n}}$, and $0 < y < 1$.

$$\Rightarrow x = \pm\sqrt{\left(\frac{1}{y}-1\right)n}.$$

Let $x_1 = +\sqrt{\left(\frac{1}{y}-1\right)n}$, $x_2 = -x_1$

Then, $$\left|\frac{dx_1}{dy}\right| = \left|\frac{\sqrt{n}}{2\cdot\sqrt{\frac{1}{y}-1}}\left(-\frac{1}{y^2}\right)\right| = \frac{\sqrt{n}}{2y^2\sqrt{\frac{1}{y}-1}} = \left|\frac{dx_2}{dy}\right|$$

The PDF of Y is,

$$f_Y(y) = \begin{cases} f_X(x_1)\left|\frac{dx_1}{dy}\right| + f_X(x_2)\left|\frac{dx_2}{dy}\right|, & \text{if } 0<y<1 \\ 0, & \text{ow} \end{cases}$$

$$= \begin{cases} \frac{(1-y)^{\frac{1}{2}-1} \cdot y^{\frac{n}{2}-1}}{\beta\left(\frac{n}{2}, \frac{1}{2}\right)}, & 0<y<1 \\ 0, & \text{ow} \end{cases}$$

Remark :–
$$Y = \frac{1}{1+\frac{X^2}{n}} = \frac{1}{1+\frac{X_1^2}{X_n^2}} = \frac{X_n^2}{X_1^2+X_n^2} \sim \text{Beta}\left(\frac{n}{2}, \frac{1}{2}\right)$$

☆ **Ex.2.** If $X, Y \overset{iid}{\sim} N(0,1)$, find the PDF of
(a) $\frac{X}{|Y|}$, (b) $\frac{X+Y}{|X-Y|}$.

Soln. →

(a) $X, Y \overset{iid}{\sim} N(0,1)$

$\Rightarrow \begin{cases} X \sim N(0,1) \\ Y^2 \sim \chi_1^2 \end{cases}$, independently

$\Rightarrow \frac{X}{\sqrt{Y^2/1}} \sim t_1$.

The PDF of $t = \frac{X}{|Y|} \sim t_1$ is

$$f_t(t) = \frac{1}{\sqrt{1}\cdot \beta\left(\frac{1}{2}, \frac{1}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{1}\right)^{\frac{1+1}{2}}}, \quad t \in \mathbb{R}$$

$$= \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R}$$

(b) $X, Y \overset{iid}{\sim} N(0,1)$

$\Rightarrow U = \frac{X+Y}{\sqrt{2}}, V = \frac{X-Y}{\sqrt{2}}$

$\overset{iid}{\sim} N(0,1)$

$\Rightarrow \frac{U}{\sqrt{V^2/1}} \sim t_1$

$\Rightarrow \frac{X+Y}{|X-Y|} \sim t_1$.

☆ **Ex.3.** Let $t \sim t_n$. Then show that $f_t(t) \longrightarrow \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}$ as $n \to \infty$

[C.U. 2009]

Hints:- $\lim_{n\to\infty} f_t(t) = \text{constant} \cdot \lim_{n\to\infty} \left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$

$$= \text{constant} \cdot \lim_{n\to\infty} \frac{\left(1+\frac{t^2}{n}\right)^{-n/2}}{\left(1+\frac{t^2}{n}\right)^{1/2}}$$

$$= \text{constant} \cdot e^{-t^2/2}, \quad t \in \mathbb{R}$$

$$= f(t), \text{ say}.$$

Here, $\int_{-\infty}^{\infty} f(t)\,dt = \text{constant} \cdot \int_{-\infty}^{\infty} e^{-t^2/2}\,dt$

$\Rightarrow \text{constant} = \frac{1}{\sqrt{2\pi}}$.

Hence, $f_t(t) \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, as $n \to \infty$.

Ex.3. **Ques:-** Obtain the limitting PDF of t-distn when $n \to \infty$.

Remark:— $\dfrac{\Gamma(n+k)}{\Gamma(n)} = \dfrac{(n+k-1)!}{(n-1)!} \simeq n^k$, for large n, by stirling's approximation.

Tails of t-distribution :—

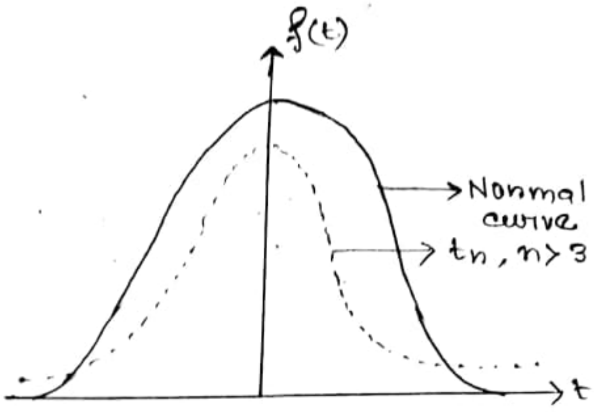
Clearly, $P[\,|t_n| > t_0\,]$

$\geq P[\,|Z| > t_0\,]$

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

$\Rightarrow$ The tails of t-distn. have a greater area than the tails of standard Normal distn..

Also, $t \sim t_n \xrightarrow{L} Z \sim N(0,1)$



Student's 't' statistic :— Let us consider sampling from $N(\mu, \sigma^2)$ population on the basis of a random sample $(X_1, \ldots, X_n)$. define,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The given μ, the statistic $\dfrac{\sqrt{n}(\bar{X}-\mu)}{s}$ is known as Student's t statistic and it follows t-distribution with $(n-1)$ degree of freedom.

Fisher's 't'- Statistic :— Let $(X_{11}, X_{12}, \ldots, X_{1n_1})$ be a random sample from $N(\mu_1, \sigma_1^2)$ and $(X_{21}, X_{22}, \ldots, X_{2n_2})$ be a random sample from $N(\mu_2, \sigma_2^2)$. Further we assume that the populations are homoscedastic, i.e. $\sigma_1^2 = \sigma_2^2$.

Let us define, $\bar{X}_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{n_i} X_{ij}$, $i = 1, 2$

and $s_i^2 = \dfrac{1}{n_i - 1}\displaystyle\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$, $i = 1, 2$.

Again, define,

$$s^2 = \frac{\sum_{j=1}^{n_1}(X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2}(X_{2j} - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 2)s_2^2}{n_1 + n_2 - 2}$$

Let, $t = \dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s\left(\sqrt{\frac{1}{n_1}} + \sqrt{\frac{1}{n_2}}\right)} \sim t_{n_1 + n_2 - 2}$

then 't' is called Fisher's 't' statistic.

C. F-distribution :

• **Definition:** Suppose X and Y are independently distributed χ^2-variables with m and n degrees of freedom. Define,

$$F = \frac{X/m}{Y/n},$$

then F is said to have a F distribution with (m, n) degrees of freedoms if the PDF is of the form —

$$f_F(F) = \begin{cases} \dfrac{\left(\frac{m}{n}\right)^{\frac{m}{2}} F^{\frac{m}{2}-1}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)\left(1+\frac{m}{n}F\right)^{\frac{m+n}{2}}} I_F(0,\infty) \\ 0 \qquad , \text{ow} \end{cases}$$

We denote, $F \sim F_{m,n}$.

Remark:- Note that, $\frac{m}{n} F = \frac{X}{Y} = \frac{\chi^2_m}{\chi^2_n} \sim$ 2nd kind Beta$\left(\frac{m}{2}, \frac{n}{2}\right)$

• **Derivation of the PDF :** [C.U. 2002]

Joint distribution of X and Y

$$f_{XY}(x,y) = \frac{1}{2^{\frac{m+n}{2}} \Gamma\frac{m}{2} \Gamma\frac{n}{2}} e^{-\frac{1}{2}(x+y)} \cdot x^{\frac{m}{2}-1} \cdot y^{\frac{n}{2}-1} I_x(0,\infty) I_y(0,\infty)$$

Let us take the transformation

$$(X, Y) \longrightarrow (F, Y).$$

$$F = \frac{Xn}{Ym}, \quad 0 < F < \infty.$$

$$\therefore X = \frac{m}{n} FY$$

$$\therefore J\left(\frac{x, y}{F, y}\right) = y \cdot \frac{m}{n}.$$

$$|J| = \left| y \cdot \frac{m}{n} \right| = \frac{ym}{n} \quad [\text{since } y > 0,\ m, n > 0]$$

The joint distribution of F, Y

$$\therefore f_{FY}(F, y) = \frac{m/n}{2^{\frac{m+n}{2}} \Gamma\frac{m}{2} \Gamma\frac{n}{2}} e^{-\frac{1}{2}\left(\frac{m}{n}Fy + y\right)} \cdot \left(\frac{m}{n}Fy\right)^{\frac{m}{2}-1} y^{\frac{n}{2}-1} \cdot y \; I_{F,y}(0,\infty)$$

$$= \frac{(F)^{\frac{m}{2}-1}}{2^{\frac{m+n}{2}} \Gamma\frac{m}{2} \Gamma\frac{n}{2}} \cdot e^{-\frac{1}{2}y\left(\frac{m}{n}F+1\right)} \cdot y^{\frac{m+n}{2}-1} \left(\frac{m}{n}\right)^{\frac{m}{2}} I_{F,y}(0,\infty).$$

$$\therefore f_F(F) = \frac{(F)^{\frac{m}{2}-1} \left(\frac{m}{n}\right)^{\frac{m}{2}}}{2^{\frac{m+n}{2}} \Gamma\frac{m}{2} \Gamma\frac{n}{2}} \int_0^\infty e^{-\frac{1}{2}y\left(\frac{m}{n}F+1\right)} \cdot y^{\frac{m+n}{2}-1}$$

$$= \frac{(F)^{\frac{m}{2}-1} \left(\frac{m}{n}\right)^{\frac{m}{2}} \Gamma\frac{m+n}{2}}{2^{\frac{m+n}{2}} \Gamma\frac{m}{2} \Gamma\frac{n}{2} \cdot \left\{\frac{1}{2}\left(1+\frac{m}{n}F\right)\right\}^{\frac{m+n}{2}}} I_F(0,\infty).$$

$$\therefore f_F(F) = \frac{(F)^{\frac{m}{2}-1}\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\beta\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}F\right)^{\frac{m+n}{2}}} I_F(0,\infty)$$

NOTE:-

i) Here $\frac{X}{Y} \sim \beta_2\left(\frac{m}{2},\frac{n}{2}\right)$. Thus F is a multiple of β variate and it posses the features of β distribution of 2nd kind.

ii) Let $F \sim F_{mn}$

Then, $mF \longrightarrow \chi^2_m$ as $n \to \infty$

iii) $F \sim F_{mn}$

$\frac{1}{F} \sim F_{nm}$

$$f_F(F) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} F^{\frac{m}{2}-1}}{\beta\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}F\right)^{\frac{m+n}{2}}} I_F(0,\infty).$$

$F' = \frac{1}{F}$

$|J| = \frac{1}{F'^2}$.

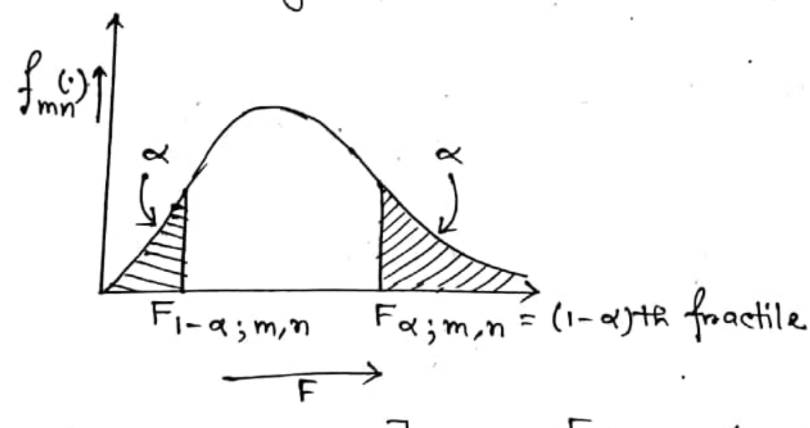
$$\therefore f_{F'}(F') = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}\left(\frac{1}{F'}\right)^{\frac{m}{2}-1}}{\beta\left(\frac{m}{2},\frac{n}{2}\right)\left(1+\frac{m}{n}\cdot\frac{1}{F'}\right)^{\frac{m+n}{2}}} \cdot \frac{1}{F'^2} I_{F'}(0,\infty)$$

$$= \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} F'^{\frac{n}{2}-1}}{\beta\left(\frac{n}{2},\frac{m}{2}\right)\left(1+\frac{n}{m}F'\right)^{\frac{m+n}{2}}} I_{F'}(0,\infty).$$

iv) $P[F_{mn} \geq c] = P[F_{n,m} \leq \frac{1}{c}]$; $c>0$

holds evidently from <iii>.

v)



$$P[F_{m,n} > F_{\alpha;m,n}] = \alpha = P\left[F_{n,m} < \frac{1}{F_{\alpha;m,n}}\right] \quad \text{(Using 4)}$$

Again,

$$P[F_{n,m} < F_{1-\alpha;n,m}] = \alpha$$

$$\therefore \frac{1}{F_{\alpha;m,n}} = F_{1-\alpha;n,m}$$

$$\therefore F_{\alpha;m,n} = \frac{1}{F_{1-\alpha;n,m}}.$$

Moments:— [C.U. 2004]

$$\mu_r' = E(F^r)$$
$$= \left(\frac{n}{m}\right)^r E\left\{\left(\frac{X}{Y}\right)^r\right\}$$
$$= \left(\frac{n}{m}\right)^r E(X^r) \cdot E(Y^{-r}).$$
$$= \left(\frac{n}{m}\right)^r \cdot \frac{2^r \Gamma\left(\frac{m}{2}+r\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot \frac{2^{-r}\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{n}{2}\right)}$$
$$; -m < 2r < n$$
$$= \left(\frac{n}{m}\right)^r \cdot \frac{\Gamma\left(\frac{m}{2}+r\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot \frac{\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{n}{2}\right)}$$

Now, $\mu_1' = \frac{n}{m} \cdot \frac{\Gamma\left(\frac{m}{2}+1\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot \frac{\Gamma\left(\frac{n}{2}-1\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{n}{\cancel{m}} \cdot \frac{\cancel{m}}{\cancel{2}} \cdot \frac{\cancel{2}}{n-2}$

$$= \frac{n}{n-2} \quad \text{if } n>2.$$

$$\mu_2' = \left(\frac{n}{m}\right)^2 \frac{\Gamma\left(\frac{m}{2}+2\right)}{\Gamma\left(\frac{m}{2}\right)} \cdot \frac{\Gamma\left(\frac{n}{2}-2\right)}{\Gamma\left(\frac{n}{2}\right)}$$
$$= \left(\frac{n}{m}\right)^2 \cdot \frac{\left(\frac{m}{2}+1\right)\left(\frac{m}{2}\right)}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)}$$
$$= \frac{n^2}{m^2} \cdot \frac{(m+2)}{\cancel{2}} \cdot \frac{m}{\cancel{2}} \cdot \frac{\cancel{2}}{(n-2)} \cdot \frac{\cancel{2}}{(n-4)}$$
$$= \frac{n^2(m+2)}{(n-2)(n-4)} \quad \text{, if } n>4.$$

$$\text{var}(F) = \mu_2' - \mu_1'^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{if } n>4.$$

Mode:— The mode of $F_{m,n}$ distribution is,

$$= \frac{n(m-2)}{m(n+2)} \quad \text{, if } m>2.$$

Skewness:— $E(F_{m,n}) = \frac{n}{n-2} > 1.$

and mode $(F_{m,n}) = \frac{n}{n+2} \cdot \frac{m-2}{m} < 1$

$\Rightarrow$ mean > mode for $F_{m,n}$ distn.

$\Rightarrow$ F-distn. is positively skewed.

☆ **Ex. 1.** If $F \sim F_{2,n}$ and $P[F > f_0] = p$, then show that

$$f_0 = \frac{n}{2}\left\{p^{-2/n} - 1\right\}.$$

Soln. →

$$p = P[F > f_0]$$

$$= \int_{f_0}^{\infty} \frac{\frac{2}{n}\left(\frac{2}{n}f\right)^{\frac{2}{2}-1}}{\beta\left(\frac{2}{2}, \frac{n}{2}\right)\left(1+\frac{2}{n}f\right)^{\frac{2+n}{2}}}\, df$$

$$= \frac{1}{\beta\left(1, \frac{n}{2}\right)} \int_{f_0}^{\infty} \frac{\frac{2}{n}\, df}{\left(1+\frac{2}{n}f\right)^{\frac{n}{2}+1}}$$

$$= \frac{1}{\beta\left(1, \frac{n}{2}\right)} \left[\frac{\left(1+\frac{2}{n}f\right)^{-\frac{n}{2}}}{\left(-\frac{n}{2}\right)}\right]_{f_0}^{\infty}$$

$$= \frac{\Gamma\left(1+\frac{n}{2}\right)}{\Gamma(1)\Gamma\left(\frac{n}{2}\right)} \times \frac{1}{\left(\frac{n}{2}\right)} \left(1+\frac{2}{n}f_0\right)^{-n/2}$$

$$= \left(1+\frac{2}{n}f_0\right)^{-n/2}$$

$$\Rightarrow f_0 = \frac{n}{2}\left\{p^{-2/n} - 1\right\}.$$

☆ **Ex. 2.** If $F \sim F_{m,n}$ distn. what will be the distn. of $\frac{1}{F}$?

Soln. → By definition, $F = \dfrac{\chi^2_m/m}{\chi^2_n/n}$

$$\Rightarrow \frac{1}{F} = \frac{\chi^2_n/n}{\chi^2_m/m} \sim F_{n,m}.$$

☆ **Ex. 3.** If ξ_p is the p^{th} quantile of $F_{m,n}$ distn. and ξ'_p the p^{th} quantile of $F_{n,m}$. Show that, $\xi_p \xi'_{1-p} = 1$.

Soln. →

$$p = P[F_{m,n} \le \xi_p]$$

$$= P\left[\frac{1}{F_{m,n}} \geq \frac{1}{\xi_p}\right]$$

$$= P\left[F_{n,m} \geq \frac{1}{\xi_p}\right]$$

$$\Rightarrow P\left[F_{n,m} \le \frac{1}{\xi_p}\right] = 1 - p.$$

But, $P\left[F_{n,m} \le \xi'_{1-p}\right] = 1-p.$

$$\Rightarrow \frac{1}{\xi_p} = \xi'_{1-p} \Rightarrow \xi_p \xi'_{1-p} = 1.$$

> If $F \sim F_{n,m}$, then $\xi_p \xi_{1-p} = 1$
>
> Then for $p = 1/2$
>
> $\xi^2_{1/2} = 1$
>
> $\Rightarrow \xi_{1/2} = \pm 1$, but as F is a positive random variable, so median of F is $+1$.

Fisher's 't' and paired 't' distribution :—

① Let $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ be a sample from $N(\mu_i, \sigma^2)$ population, $i = 1, 2$.

(The two populations are homoscedastic)

Define, $$\overline{X_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$$

$$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \overline{X_i})^2$$

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

We know, $\overline{X_i} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right)$

and, $\frac{(n_i-1)s_i^2}{\sigma^2} \sim \chi^2_{n-1}$.

They are independent, and also independent for each i.

$$\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$$

$$\therefore \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1) \text{ ------- ①}$$

$$\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

$$\frac{(n_1+n_2-2)s^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2} \text{ -------- ②}$$

$\therefore$ From ① & ② as they are independently distributed.

$$\frac{(\overline{X}_1 - \mu_1) - (\overline{X}_2 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Bigg/ \sqrt{\frac{(n_1+n_2-2)s^2}{\sigma^2}\Big/ n_1+n_2-2} \sim t_{n_1+n_2-2}$$

i.e. $$\frac{(\overline{X}_1 - \mu_1) - (\overline{X}_2 - \mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Given, $\mu_2 - \mu_2$, $\frac{(\overline{X}_1 - \mu_1) - (\overline{X}_2 - \mu_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

is termed as Fisher's 't' statistic and its sampling distn is known as Fisher's 't' distribution. Fisher's 't' statistic is usually adopted for testing the equality or difference of means of two univariate normal homoscedastic distribution.

Scanned by CamScanner

(2). Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random variable from (Bivariate Normal) $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ population. Define,

$Z_i = X_i - Y_i$ as (X_i, Y_i)'s are iid so, the Z_i's are iid.

Let, $\mu_Z = E(Z_1)$ and $\sigma_Z^2 = V(Z_1)$

then, Z_i's are iid $N(\mu_Z, \sigma_Z^2)$.

Define, $\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i$

$$\mu_Z = E(Z_1) = E(X_1 - Y_1) = \mu_x - \mu_y$$

$$\sigma_Z^2 = V(Z_1) = V(X_1 - Y_1) = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

$$\therefore \frac{\sqrt{n}(\bar{Z} - \mu_Z)}{\sigma_Z} \sim N(0,1)$$

Define, $s_Z^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \{(X_i - \bar{X}) - (Y_i - \bar{Y})\}^2$

$= s_x^2 + s_y^2 - 2r s_x s_y$,

where, $s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$r = \frac{s_{xy}}{s_x s_y}.$$

Now, $\frac{(n-1)s_Z^2}{\sigma_Z^2} \sim \chi^2_{n-1}$, s is independent of $\bar{Z}$.

$$\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{\sigma_Z} \Big/ \sqrt{\frac{(n-1)s^2}{\sigma_Z^2} \Big/ n-1} = \frac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z} \sim t_{n-1}.$$

Given μ_Z.

$\frac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z}$ is termed as Paired 't' statistic, where, sampling distribution is t_{n-1}, Paired 't' statistic is usually adopted to test the equality on difference of the means involved in a Bivariate normal population.

x ——— x

DISTRIBUTION FUCTIONS OF RANDOM VARIABLES

$\Rightarrow$ Case of Discrete variables: — If $X_1, X_2, \ldots, X_n$ are jointly distributed random variables with joint p.m.f. $f(x_1, x_2, \ldots, x_n)$ and if $Y = g(X_1, X_2, \ldots, X_n)$ is a function of $X_1, X_2, \ldots, X_n$, then the p.m.f. of Y is given by,

$$f_Y(y) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n).$$
$$\{(x_1, x_2, \ldots, x_n) : g(x_1, x_2, \ldots, x_n) = y\}$$

Example: — 1. Let $X_1 \sim Bin(m_1, p)$, $X_2 \sim Bin(m_2, p)$.
$X_1 + X_2 \sim ?$

$\Rightarrow$ The joint PMF of X_1, X_2 is,

$$f(x_1, x_2) = \binom{m_1}{x_1} p^{x_1} (1-p)^{m_1 - x_1} \binom{m_2}{x_2} p^{x_2} (1-p)^{m_2 - x_2}$$

Let us consider a function of X_1 and X_2 given by,

$$Y = X_1 + X_2.$$

Then the PMF of Y is,

$$P[Y = y] = f_Y(y) = \sum_{x_1=0}^{m_1} \sum_{x_2=0}^{m_2} f(x_1, x_2)$$
$$\{(x_1, x_2) : x_1 + x_2 = y\}$$

$$= \sum_{x_1=0}^{m_1} \sum_{x_2=0}^{m_2} \binom{m_1}{x_1} \binom{m_2}{x_2} p^{x_1 + x_2} (1-p)^{m_1 + m_2 - \overline{x_1 + x_2}}$$
$$\{(x_1, x_2) : x_1 + x_2 = y\}.$$

$$= \sum_{x_1=0}^{m_1} \sum_{x_2=0}^{m_2} \binom{m_1}{x_1} \binom{m_2}{x_2} p^{y} (1-p)^{m_1 + m_2 - y}$$
$$\{(x_1, x_2) : x_1 + x_2 = y\}$$

$$= p^y (1-p)^{m_1 + m_2 - y} \sum_{x_1=0}^{m_1} \sum_{x_2=0}^{m_2} \binom{m_1}{x_1} \binom{m_2}{x_2}$$
$$\{(x_1, x_2) : x_1 + x_2 = y\}$$

$$= p^y (1-p)^{m_1 + m_2 - y} \sum_{x_1=0}^{m_1} \binom{m_1}{x_1} \binom{m_2}{y - x_1}$$

Now, $\binom{m_1}{x_1}$ = coefficient of t^{x_1} in $(1+t)^{m_1}$

$\binom{m_2}{y - x_1}$ = coefficient of $t^{y - x_1}$ in $(1+t)^{m_2}$

$\therefore \sum_{x_1=0}^{m_1} \binom{m_1}{x_1} \binom{m_2}{y - x_1}$ = coefficient of t^y in $(1+t)^{m_1 + m_2} = \binom{m_1 + m_2}{y}$

i.e. $f_Y(y) = \binom{m_1 + m_2}{y} p^y (1-p)^{m_1 + m_2 - y}$, i.e. $Y = X_1 + X_2 \sim Bin(m_1 + m_2, p)$.

2. <u>Example</u> :- Let X_1 and X_2 be two jointly distributed poisson random variable, then find the distn of X_1+X_2.

$\Rightarrow$ The joint PMF is $f(x_1,x_2) = \frac{e^{-\lambda_1} \cdot \lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2} \cdot \lambda_2^{x_2}}{x_2!}$, $x_1, x_2 = 0,1,2,..\infty$

Let $Y = X_1+X_2$ be a function of X_1 and X_2, then the PMF of Y is,

$$P[Y=y] = f_Y(y) = \sum_{\substack{x_1=0 \\ \{(x_1,x_2):x_1+x_2=y\}}}^{\infty} \sum_{x_2=0}^{\infty} e^{-(\lambda_1+\lambda_2)} \cdot \frac{\lambda_1^{x_1} \cdot \lambda_2^{x_2}}{x_1!\, x_2!}$$

$$= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{y!} \sum_{x_1=0}^{\infty} \frac{y!}{(y-x_1)!\, x_1!} \cdot \lambda_1^{x} \lambda_2^{y-x_1}$$

$$= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{y!} \sum_{x_1=0}^{\infty} \binom{y}{x_1} \lambda_1^{x_1} \lambda_2^{y-x_1}$$

$$= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{y!} (\lambda_1+\lambda_2)^y \quad , \quad y = 0,1,2,\ldots$$

$\therefore Y \sim$ Poisson $(\lambda_1+\lambda_2)$.

3. If $X_1 \sim Bin(m_1, p)$ and $X_2 \sim Bin(m_2, p)$ and if X_1, and X_2 are independent, find the distribution of X_1 given that $X_1+X_2 = t$

$$\Rightarrow P[X = x_1 \mid X_1+X_2 = t] = \frac{P[X_1 = x_1, X_1+X_2 = t]}{P[X_1+X_2 = t]}$$

$$= \frac{P[X_1 = x_1, X_2 = t-x_1]}{P[X_1+X_2 = t]}$$

$$= \frac{P[X_1 = x_1]\, P[X_2 = t-x_1]}{P[X_1+X_2 = t]}$$

$$= \frac{\binom{m_1}{x_1} p^{x_1} (1-p)^{m_1-x_1} \binom{m_2}{t-x_1} p^{t-x_1} (1-p)^{m_2-t+x_1}}{\binom{m_1+m_2}{t} p^t (1-p)^{m_1+m_2-t}}$$

$$= \frac{\binom{m_1}{x_1}\binom{m_2}{t-x_1}}{\binom{m_1+m_2}{t}}.$$

4. $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$ and if X_1 and X_2 are independent, then $X_1 / X_1 + X_2 = t \sim ?$

$\Rightarrow$

$$P[X_1 = x_1 / X_1 + X_2 = t] = \frac{P[X_1 = x_1, X_1 + X_2 = t]}{P[X_1 + X_2 = t]}$$

$$= \frac{P[X_1 = x_1, X_2 = t - x_1]}{P[X_1 + X_2 = t]}$$

$$= \frac{P[X_1 = x_1]\, P[X_2 = t - x_1]}{P[X_1 + X_2 = t]}$$

$$= \frac{e^{-\lambda_1} \cdot \frac{\lambda_1^{x_1}}{x_1!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_2^{t-x_1}}{(t-x_1)!}}{e^{-(\lambda_1+\lambda_2)} \cdot \frac{(\lambda_1+\lambda_2)^t}{t!}}$$

$$= \binom{t}{x_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x_1} \left(1 - \frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{t-x_1},$$

$$x_1 = 0, 1, 2, \ldots, t$$

which is a binomial probability with success probability, $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$.

5. $X \sim \text{Bin}(m, \theta)$; $2X \sim ?$

$\Rightarrow$ $X \sim \text{Bin}(m, \theta)$, then the PMF is,

$$P[X = x] = \binom{m}{x} \theta^x (1-\theta)^{m-x} \; ; \; x = 0, 1, \ldots, m.$$

$Y = 2X$, $\quad y = 0, 2, 4, \ldots, 2m$.

$$P[Y = y] = P\left[X = \frac{y}{2}\right]$$

$$= \binom{m}{y/2} \theta^{y/2} (1-\theta)^{m-y/2}, \quad y = 0, 2, 4, \ldots, 2m$$

$\therefore Y \sim \text{Bin}(m, \theta)$.

2) <u>Case of Continuous variables:</u> — Let X be a continuous random variable with distribution function $F_X(x)$. Also let $Y = g(X)$ be a function of X then the distribution function of Y is

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

Let us consider that $Y = g(X)$ is a monotonic function of X either monotonically increasing or, monotonically decreasing.

If $g(x)$ is increasing then $g(X) \le y \Leftrightarrow X \le g^{-1}(y)$

again if $g(x)$ is decrasing then $g(X) \le y \Leftrightarrow X \ge g^{-1}(y)$.

$$\therefore F_Y(y) = P[g(X) \le y] = \begin{cases} P[X \le g^{-1}(y)] & \text{if } g(x) \text{ is increasing.} \\ P[X \ge g^{-1}(y)] & \text{if } g(x) \text{ is decreasing.} \end{cases}$$

$$\text{i.e. } F_Y(y) = \begin{cases} F_X[g^{-1}(y)] & \text{when } g(x) \text{ is increasing,} \\ 1 - F_X[g^{-1}(y)] & \text{when } g(x) \text{ is decreasing.} \end{cases}$$

If we denote f_X and f_Y, the PDF of X and Y, respectively, then

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \dfrac{d g^{-1}(y)}{dy}, & \text{if } g(x) \text{ is increasing.} \\ -f_X[g^{-1}(y)] \dfrac{d g^{-1}(y)}{dy}, & \text{if } g(x) \text{ is decreasing.} \end{cases} \qquad (*)$$

Now, note that,

$g(\cdot)$ is increasing or decreasing $\Leftrightarrow$ $g^{-1}(\cdot)$ is increasing or decreasing

and

$\dfrac{d g^{-1}(y)}{dy}$ is positive or negative according as $g(x)$ is increasing or decreasing, respectively.

Thus it appears from $(*)$ is,

$$\therefore \boxed{f_Y(y) = f_X[g^{-1}(y)] \left| \frac{d g^{-1}(y)}{dy} \right|}$$

Example: – 1. Let $X \sim N(\mu, \sigma^2)$, find the PDF of $Y = a + bX$.

$\Rightarrow$ the PDF of X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{ where } -\infty < x < \infty.$$

If $f_Y(y)$ is the PDF of $Y = a + bX$ then,

$$f_Y(y) = f_X\left(\frac{y-a}{b}\right) \left| \frac{d}{dy}\left(\frac{y-a}{b}\right) \right|$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{y-a}{b} - \mu}{\sigma}\right)^2} \cdot \frac{1}{|b|}$$

$$= \frac{1}{\sigma|b|\sqrt{2\pi}} e^{-\frac{1}{2b^2\sigma^2}(y-a-b\mu)^2}$$

which is the PDF of Normal distribution with mean $a + b\mu$ and variance $b^2\sigma^2$.

2. $X \sim N(0,1)$, find the distn of $Y = e^X$.

$\Rightarrow$ $Y = g(X) \longrightarrow f_Y(y) = f_X[g^{-1}(y)] \left| \frac{dg^{-1}(y)}{dy} \right|$

PDF of X is,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

If $f_Y(y)$ is the PDF of $Y = e^X \Rightarrow X = \ln Y$

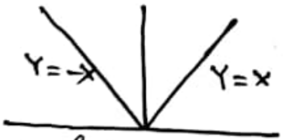
then $f_Y(y) = f_X(\log_e Y) \left| \frac{d \log_e y}{dy} \right|$

$$= \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}(\log_e y)^2}, \quad 0 < y < \infty.$$

Here $Y = e^X$ is called a lognormal variable.

3. $X \sim N(0,1)$, find the PDF of $Y = |X|$.

$\Rightarrow$



$$Y = \begin{cases} X & \text{if } X > 0 \\ 0 & \text{if } X = 0 \\ -X & \text{if } X < 0 \end{cases}$$

the PDF of X is, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty$

Let us consider the transformation,

$Y \longrightarrow X$ such that $Y = |X|$, i.e. $X = \begin{cases} Y & \text{if } 0 < X < \infty \\ -Y & \text{if } -\infty < X < 0 \end{cases}$

For both the inverse transformations of the transformation, the Jacobian is $|J| = 1$.

So, the PDF of $Y=|X|$ is

$$f_Y(y) = [f_X(y) + f_X(-y)]$$

$$= \frac{1}{\sqrt{2\pi}} e^{-y^2/2} + \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$= \sqrt{\frac{2}{\pi}} e^{-y^2/2} \quad , 0<y<\infty$$

4. $X \sim N(0,1)$, find the distn. of $Y=X^2$.

⇒ Let us consider the transformation $Y=X^2$ so that

$$X = \begin{cases} \sqrt{Y} & \text{if } 0<X<\infty \\ -\sqrt{Y} & \text{if } -\infty<X<0 \end{cases}$$

For both the cases, the jacobian of the transformation is,

$$|J| = \frac{1}{2\sqrt{y}}$$

So, the PDF of $Y=X^2$ is,

$$f_Y(y) = \left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right] \cdot \frac{1}{2\sqrt{y}}$$

$$= \left[\frac{1}{\sqrt{2\pi}} e^{-y/2} + \frac{1}{\sqrt{2\pi}} e^{-y/2}\right] \cdot \frac{1}{2\sqrt{y}}$$

$$= \sqrt{\frac{2}{\pi}} e^{-y/2} \cdot y^{1/2-1}$$

$$= \frac{1}{2^{-1/2}} \cdot \frac{e^{-y/2} \cdot y^{1/2-1}}{\Gamma(\frac{1}{2})} \quad , 0<y<\infty$$

Note that,

$$\int_0^\infty f_Y(y)\,dy = \int_0^\infty \frac{1}{2^{-1/2}} \cdot \frac{1}{\Gamma(\frac{1}{2})} e^{-y/2} \cdot \left(\frac{y}{2}\right)^{1/2-1} \cdot 2^{-1/2}\,dy$$

$$= \frac{\int_0^\infty e^{-y/2} (y/2)^{1/2-1}\,dy}{\Gamma(\frac{1}{2})}$$

$$= \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2})} = 1.$$

Case I :- Let $X_1, X_2, \ldots, X_n$ is a set of n jointly distributed random variables with joint PDF $f_X(x_1, x_2, \ldots, x_n)$; we consider the transformation

$$(X_1, X_2, \ldots, X_n) \longrightarrow (Y_1, Y_2, \ldots, Y_n)$$

such that

$$Y_1 = g_1(X_1, \ldots, X_n)$$
$$Y_2 = g_2(X_1, \ldots, X_n)$$
$$\vdots$$
$$Y_n = g_n(X_1, \ldots, X_n)$$

Assuming that the inverse transformation exists

$$X_1 = \varphi_1(Y_1, \ldots, Y_n)$$
$$X_2 = \varphi_2(Y_1, \ldots, Y_n)$$
$$\vdots$$
$$X_n = \varphi_n(Y_1, \ldots, Y_n)$$

In that case the joint PDF of $Y_1, \ldots, Y_n$ is given by,

$$f_Y(y_1, \ldots, y_n) = f_X[\varphi_1(y), \varphi_2(y), \ldots, \varphi_n(y)]\,|J|.$$

where, $|J| = \left| J\left(\frac{x_1, x_2, \ldots, x_n}{y_1, y_2, \ldots, y_n}\right)\right|$

$$= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

$$= \left| \frac{\partial(\text{old variable})}{\partial(\text{new variable})} \right|$$

Example. 1. $X_1 \sim \text{Gamma}(\alpha, p_1)$, $X_2 \sim \text{Gamma}(\alpha, p_2)$, find the distribution of $X_1 + X_2$.

Ans :-

Joint PDF of X_1 & X_2 is

$$f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma(p_1)\Gamma(p_2)} \cdot e^{-\alpha(x_1+x_2)} \cdot x_1^{p_1-1} \cdot x_2^{p_2-1}.$$

$$X_1, X_2 \longrightarrow Y, Z$$

$Y = X_1 + X_2$, $0 < Y < \infty$

$Z = X_2 = Y - X_1$, $0 < Z < Y$

The Jacobian tramformation is

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$$

so, the joint PDF of Y, Z is

$$f_{YZ}(y, z) = \frac{\alpha^{p_1+p_2}}{\Gamma(p_1)\Gamma(p_2)} \cdot e^{-\alpha y}(y-z)^{p_1-1} \cdot z^{p_2-1}$$

So, the marginal PDF of Y is, $0<y<\infty$, $0<z<y$

$$f_Y(y) = \int_0^y f_{YZ}(y, z)\,dz$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma(p_1)\Gamma(p_2)} \cdot e^{-\alpha y} \int_0^y (y-z)^{p_1-1} z^{p_2-1}\,dz$$

$$= \frac{\alpha^{p_1+p_2} e^{-\alpha y}}{\Gamma(p_1)\Gamma(p_2)} \cdot I$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma(p_1+p_2)} e^{-\alpha y} y^{p_1+p_2-1}, \quad 0<y<\infty$$

$\therefore Y = X_1 + X_2 \sim \text{Gamma}(\alpha, p_1+p_2)$.

Let, $I = \int_0^y (y-z)^{p_1-1} \cdot z^{p_2-1}\,dz$

Let, $u = \frac{z}{y}$, $du = \frac{1}{y}dz$, $0<u<1$

$$\therefore I = \int_0^1 [(1-u)y]^{p_1-1} (uy)^{p_2-1} y\,du$$

$$= y^{p_1+p_2-1} \int_0^1 (1-u)^{p_1-1} u^{p_2-1}\,du$$

$$= y^{p_1+p_2-1}\, B(p_1, p_2)$$

$$= y^{p_1+p_2-1} \cdot \frac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}.$$

2. $X \sim G(\alpha, p_1)$ & $Y \sim G(\alpha, p_2)$, then find the distribution of $U = X+Y$ and $V = \frac{X}{X+Y}$. [C.U. 2010]

ANS:— Let us start from the joint PDF of X and Y given by,

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} \cdot e^{-\alpha(x+y)} \cdot x^{p_1-1} \cdot y^{p_2-1}$$

now, let us make the transformation $(X,Y) \longrightarrow (U,V)$

$$\left.\begin{array}{l} U = X+Y \\ V = \frac{X}{X+Y} \end{array}\right\} \begin{array}{l} X = UV \\ Y = U(1-V) \end{array} \qquad \begin{array}{l} 0<U<\infty \\ 0<V<1 \end{array}$$

<u>Jacobian of Transformation</u>:—

$$J = J\left(\frac{X,Y}{U,V}\right) = \begin{vmatrix} V & U \\ 1-V & -U \end{vmatrix} = -U$$

so, $|J| = U$.

Thus, the joint PDF of U and V is,

$$f_{UV}(u,v) = \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} \cdot e^{-\alpha u} (uv)^{p_1-1} [u(1-v)]^{p_2-1} \cdot u$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} e^{-\alpha u} \cdot u^{p_1+p_2-1} \cdot v^{p_1-1} (1-v)^{p_2-1}, \quad \begin{array}{l} 0<u<\infty \\ 0<v<1 \end{array}$$

so, the marginal PDF of U is,

$$f_U(u) = \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} e^{-\alpha u} \cdot u^{p_1+p_2-1} \int_0^1 v^{p_1-1} (1-v)^{p_2-1} dv$$

$$= \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} \cdot e^{-\alpha u} \cdot u^{p_1+p_2-1} \cdot \beta(p_1+p_2)$$

$$= \frac{\alpha^{p_1+p_2} \cdot e^{-\alpha u} \cdot u^{p_1+p_2-1}}{\Gamma p_1+p_2}, \quad 0<u<\infty$$

Marginal PDF of V is,

$$f_V(v) = \frac{\alpha^{p_1+p_2}}{\Gamma p_1 \Gamma p_2} \cdot v^{p_1-1} (1-v)^{p_2-1} \int_0^\infty e^{-\alpha u} u^{p_1+p_2-1} du$$

$$= \frac{1}{\beta(p_1,p_2)} v^{p_1-1} (1-v)^{p_2-1}, \quad 0<v<1$$

$\therefore V \sim \beta(p_1, p_2)$.

Distribution of the linear function of independent normal variates :

Ques:- Let X_i $(i=1,2,\ldots,n)$ be n independent random variables such that

$$X_i \sim N(\mu_i, \sigma_i^2).$$

Find the distribution of $Y = a + \sum_{i=1}^{n} b_i X_i$ [C.U. 2010]

Ans:- The joint PDF of $(X_1, X_2, \ldots, X_n)$ is

$$f_{X_1,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \left(\frac{1}{2\pi}\right)^{n/2} \cdot \frac{1}{\prod_{i=1}^{n}\sigma_i} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right].$$

Let us take the transformation,

$$(X_1,\ldots,X_n) \longrightarrow (Y_1,\ldots,Y_n).$$

$$Y_1 = \frac{b_1\sigma_1\left(\frac{X_1-\mu_1}{\sigma_1}\right) + \cdots + b_n\sigma_n\left(\frac{X_n-\mu_n}{\sigma_n}\right)}{\sqrt{\sum b_i^2\sigma_i^2}}$$

$$\vdots$$

$$Y_j = l_{j1}\left(\frac{X_1-\mu_1}{\sigma_1}\right) + \cdots + l_{jn}\left(\frac{X_n-\mu_n}{\sigma_n}\right), \quad j = 2(1)n.$$

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{bmatrix} \frac{b_1\sigma_1}{\sqrt{\sum b_i^2\sigma_i^2}} & \cdots & \frac{b_n\sigma_n}{\sqrt{\sum b_i^2\sigma_i^2}} \\ l_{21} & \cdots & l_{2n} \\ \vdots & & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{pmatrix} \frac{X_1-\mu_1}{\sigma_1} \\ \vdots \\ \frac{X_n-\mu_n}{\sigma_n} \end{pmatrix}$$

when l_{jk}'s are so chosen that the whole transformation is orthogonal.

or, $\underset{\sim}{Y} = L\underset{\sim}{W}$

$$\therefore \underset{\sim}{Y}'\underset{\sim}{Y} = \underset{\sim}{W}'L'L\underset{\sim}{W}$$

$$\Rightarrow \sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} W_i^2 \quad [\because L'L = I]$$

$$= \sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$$

Jacobian is,

$$J = \left| J\left(\frac{X_1, X_2, \ldots, X_n}{Y_1, Y_2, \ldots, Y_n}\right)\right|$$

$$= \frac{1}{\left| J\left(\frac{Y_1, Y_2, \ldots, Y_n}{X_1, X_2, \ldots, X_n}\right)\right|}$$

$$\therefore J = \cfrac{1}{\begin{vmatrix} \frac{1}{\sigma_1} \times \frac{b_1\sigma_1}{\sqrt{\sum b_i^2\sigma_i^2}} & \cdots & \frac{1}{\sigma_n} \cdot \frac{b_n\sigma_n}{\sqrt{\sum b_i^2\sigma_i^2}} \\ \frac{1}{\sigma_1} \times l_{21} & \cdots & \frac{1}{\sigma_n} l_{2n} \\ \vdots & & \vdots \\ \frac{1}{\sigma_1} \times l_{n1} & \cdots & \frac{1}{\sigma_n} l_{nn} \end{vmatrix}}$$

$$= \frac{\prod_{i=1}^{n} \sigma_i}{|L|} = \pm \prod_{i=1}^{n} \sigma_i$$

$$\therefore |J| = \prod_{i=1}^{n} \sigma_i$$

So, the joint PDF of $Y_1, Y_2, \ldots, Y_n$ is

$$h(y_1, y_2, \ldots, y_n) = \left(\frac{1}{2\pi}\right)^{n/2} \cdot \frac{1}{\prod_{i=1}^{n} \sigma_i} \prod_{i=1}^{n} \sigma_i \exp\left[-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right]$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-y_i^2/2}$$

thus $Y_1, Y_2, \ldots, Y_n$ are independent & $Y_i \sim N(0,1)$, for $i=1(1)n$.

$$Y_1 = \frac{\sum_i b_i\sigma_i \left(\frac{X_i - \mu_i}{\sigma_i}\right)}{\left(\sqrt{\sum b_i^2\sigma_i^2}\right)}$$

$$\therefore \sum_{i=1}^{n} b_i X_i = \sum_{i=1}^{n} b_i \mu_i + \sqrt{\sum_{i=1}^{n} b_i^2 \sigma_i^2}\, Y_1$$

$$\Rightarrow a + \sum_{i=1}^{n} b_i X_i = a + \sum_{i=1}^{n} b_i \mu_i + \sqrt{\sum_{i=1}^{n} b_i^2 \sigma_i^2}\, Y_1$$

$$= A + BY_1$$

where $Y_1 \sim N(0,1)$

$$\therefore a + \sum_{i=1}^{n} b_i X_i \sim N(A, B^2)$$

Sampling Distribution of a Statistic.

If let $X_1, X_2, \dots, X_n$ be a random sample of size n dracon on a random variable X.
$X_1, X_2, \dots, X_n$ are iid random variables with common distributn function $F_\theta(x)$. A function of $X_1, X_2, \dots, X_n$ is given by,

$T = T(X_1, X_2, \dots, X_n)$ is called a statistic and the probability distribution of X is called the sampling distribution of T.

Chi-square Statistic & its distribution :-

Let $(X_1, X_2, \dots, X_n)$ is a random sample on a random variable X where, $X \sim N(\mu, \sigma^2)$, then the statistic is defined by,

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$ is called a chi-square statistic with degree of freedom n.

WLG we can take $\mu = 1, \sigma = 1$, so that

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

i.e. in any case chi-square statistic with n degree of freedom is the sum of square of n independent standard normal variables.

Distribution of χ^2-statistic: To find the distribution of χ^2, we start from the joint PDF of $X_1, X_2, \dots, X_n$ given by $f(x_1, x_2, \dots, x_n)$

$$f(x_1, x_2, \dots, x_n) = \left(\frac{1}{2\pi} \right)^{n/2} e^{-\frac{1}{2} \sum_{i=1}^{n} x_i^2}, \quad -\infty < x_i < \infty \;\; \forall\, i = 1(1)n.$$

Note, $$\beta(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1}\, dx$$

Putting $x = \sin^2\theta$
$dx = 2\sin\theta\cos\theta\, d\theta$

$$= 2 \int_0^{\pi/2} \sin^{2p-1}\theta \cos^{2q-1}\theta\, d\theta$$

or, $$\beta\left(\frac{p+1}{2}, \frac{q+1}{2} \right) = 2 \int_0^{\pi/2} \sin^p\theta \cos^q\theta\, d\theta$$

Now let us take the tranformation,

$$(X_1, X_2, \dots, X_n) \longrightarrow (\chi, \theta_1, \theta_2, \dots, \theta_{n-1})$$

such that,

$X_1 = \chi \cos\theta_1$
$X_2 = \chi \sin\theta_1 \cos\theta_2$
$X_3 = \chi \sin\theta_1 \sin\theta_2 \cos\theta_3$
$\vdots$
$X_{n-1} = \chi \sin\theta_1 \sin\theta_2 \cdots \sin\theta_{n-2} \cos\theta_{n-1}$
$X_n = \chi \sin\theta_1 \cdots\cdots \sin\theta_{n-1}$.

$0 < \chi < \infty$
$0 < \theta_i < \pi \quad \forall \; i = 1(1)n-2$
$0 < \theta_{n-1} < 2\pi$

for this tramformation, $\sum_{i=1}^{n} X_i^2 = \chi^2$
and jacobian of the tramformation is

$$J = \chi^{n-1} (\cos\theta_1)^{n-2} (\cos\theta_2)^{n-3} \cdots (\cos\theta_{n-2})$$

so, the joint PDF of $\chi, \theta_1, \theta_2, \ldots, \theta_{n-1}$ is

$f(\chi, \theta_1, \theta_2, \ldots, \theta_{n-1})$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\chi^2\right] \chi^{n-1} (\sin\theta_1)^{n-2} (\sin\theta_2)^{n-3} \cdots (\sin\theta_{n-2})$$

so, if $f_\chi(y)$ be the marginal PDF of χ then,

$$f_\chi(y) = \left(\frac{1}{2\pi}\right)^{n/2} y^{n-1} e^{-y^2/2} \cdot 2\pi \cdot \beta\left(1, \tfrac{1}{2}\right) \beta\left(\tfrac{3}{2}, \tfrac{1}{2}\right) \cdots \beta\left(\tfrac{n-2}{2}, \tfrac{1}{2}\right) \beta\left(\tfrac{n-1}{2}, \tfrac{1}{2}\right)$$

$$= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}-1} y^{n-1} e^{-y^2/2} \cdot \frac{(\sqrt{\pi})^{n-2}}{\Gamma n/2}$$

$$= \frac{e^{-y^2/2} \, y^{n-1}}{2^{n/2-1} \, \Gamma n/2}$$

if $z = \chi^2$
so, the PDF of χ^2 is the above,
We take the tramformation,

$z = y^2$

$\frac{dz}{dy} = 2y$

$|J| = \frac{1}{2\sqrt{z}}$

$$\therefore f_{\chi^2}(z) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-z/2} z^{\frac{n-1}{2}} \cdot \frac{1}{\sqrt{z}}$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-z/2} \cdot z^{n/2-1}, \quad 0<z<\infty$$

Note that if $f_{\chi^2}(z) \geq 0$

$$\int_0^\infty f_{\chi^2}(z)\,dz = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{-z/2} z^{n/2-1}\,dz = 1.$$

Calculation of Mean & Variance: —

$X_1, X_2, \ldots, X_n \longrightarrow$ independent random variable

$X_i \sim N(0,1)$

$$\chi^2 = \sum_{i=1}^n X_i^2$$

$$f_{\chi^2}(z) = \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-z/2} z^{\frac{n}{2}-1}, \quad 0<z<\infty$$

$$E(\chi^2)^m = \int_0^\infty \frac{z^m}{2^{n/2}\Gamma\frac{n}{2}} e^{-z/2} \cdot z^{\frac{n}{2}-1}\,dz$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \int_0^\infty e^{-z/2} z^{\frac{n+2m}{2}-1}\,dz$$

$$= \frac{1}{2^{n/2}\Gamma\frac{n}{2}} \cdot 2^{\frac{n+2m}{2}} \cdot \Gamma\left(\frac{n+2m}{2}\right)$$

$$\therefore E(\chi^2) = n, \quad V(\chi^2) = 2n.$$

Joint distribution of the sample mean and sample variance: —

Result: Let the sample is drawn from a normal population. Let $(X_1, X_2, \ldots, X_n)$ be a random sample drawn from normal population $N(\mu, \sigma^2)$, let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

be the sample mean and the sample variance. Then

i) $\bar{X}$ & s^2 are independently distributed,

ii) $\bar{X} \sim N(\mu, \sigma^2/n)$ & $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. [C.U. 2005]

Proof:— To prove the above result let us start from the joint PDF of $X_1, X_2, \ldots, X_n$ given by,

$$f_X(x_1, x_2, \ldots, x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \quad ; -\infty < x_i < \infty$$

Now, let us consider the transformation

$$(X_1, X_2, \ldots, X_n) \longrightarrow (Y_1, Y_2, \ldots, Y_n) \text{ such that}$$

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n)$$

$$Y_j = l_{j1}X_1 + l_{j2}X_2 + \cdots + l_{jn}X_n \quad \text{for } j = 2(1)n.$$

where l_{jk}'s are chosen so that the whole transformation is orthogonal.

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ l_{21} & \cdots & l_{2n} \\ \vdots & & \vdots \\ l_{n1} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

---

for e.g. one such choice is,

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$$

$$Y_2 = \frac{1}{\sqrt{n(n-1)}}(X_1 + X_2 + \cdots + X_{n-1} - \overline{n-1}\,X_n)$$

$$\vdots$$

$$Y_{n-1} = \frac{1}{\sqrt{3.2}}(X_1 + X_2 - 2X_3)$$

$$Y_n = \frac{1}{2}(X_1 - X_2)$$

Then under the transformation

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ l_{21} & \cdots & l_{2n} \\ \vdots & & \vdots \\ l_{n1} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = L\underset{\sim}{X}$$

The jacobian of the transformation is $|L| = 1$.

$$Y'Y = \sum_{i=1}^{n} Y_i^2 = (L\underset{\sim}{X})'(L\underset{\sim}{X})$$

$$= \underset{\sim}{X}'\underset{\sim}{X} = \sum_{i=1}^{n} X_i^2$$

$$f_Y(y_1, y_2, \ldots, y_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_i^2 + n\mu^2 - 2\mu\sqrt{n}\, y_1\right)}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\left\{(y_1 - \sqrt{n}\mu)^2 + \sum_{i=1}^{n} y_i^2\right\}}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) e^{-\frac{1}{2\sigma^2}(y_1 - \sqrt{n}\mu)^2} \cdot \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-y_i^2/2\sigma^2}$$

i.e. $Y_1, \ldots, Y_n$ are independent.

since, $\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i^2 - 2\mu x_i + \mu^2)$

$$= \sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2$$

$$= \sum_{i=1}^{n} y_i^2 - 2\mu\sqrt{n}\, y_1 + n\mu^2$$

Here, $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$, $Y_i \sim N(0, \sigma^2)$ [where $i = 2(1)n$]

Now, $Y_1 = \sqrt{n}\,\bar{X}$.

$$\therefore \bar{X} = \frac{1}{\sqrt{n}} Y_1$$

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Also,

$$\frac{1}{\sigma^2} \sum_{i=2}^{n} Y_i^2 = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} Y_i^2 - Y_1^2\right]$$

$$= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2}$$

$$= \frac{(n-1)s^2}{\sigma^2}.$$

i.e. $\frac{(n-1)s^2}{\sigma^2} = \sum_{i=2}^{n} \left(\frac{Y_i}{\sigma}\right)^2$ = sum of squares of $(n-1)$ standard normal variates and hence follows χ^2_{n-1}.

Scanned by CamScanner

☆ **Problem:—** Let $X_1, X_2, \ldots, X_n$ be independently distributed r.v.s such that, $X_i \sim N(\mu, \sigma_i^2)\ \forall\ i=1(1)n$.

let, $\omega_i = \frac{1}{\sigma_i^2}$ and $\overline{X}_\omega = \sum_{i=1}^{n} \omega_i x_i \Big/ \sum_{i=1}^{n} \omega_i$

and, $s_\omega^2 = \sum_{i=1}^{n} \omega_i (x_i - \overline{x}_\omega)^2$

Then S.T. $\overline{X}_\omega$ and s_ω^2 are independently distributed and also find their distribution.

Solution:— Let $Y_i = \frac{X_i - \mu}{\sigma_i} = \sqrt{\omega_i}\,(X_i - \mu)\ \forall\ i=1(1)n$

then the joint PDF of $Y_1, \ldots, Y_n$ is

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2}, \quad -\infty < y_i < \infty$$

Let us consider the tramformation,

$$(Y_1, \ldots, Y_n) \longrightarrow (Z_1, Z_2, \ldots, Z_n)$$

$$Z_1 = \frac{\sqrt{\omega_1}}{\sqrt{\sum \omega_i}} Y_1 + \frac{\sqrt{\omega_2}}{\sqrt{\sum \omega_i}} Y_2 + \cdots + \frac{\sqrt{\omega_n}}{\sqrt{\sum \omega_i}} Y_n$$

$$Z_2 = l_{21} Y_1 + l_{22} Y_2 + \cdots + l_{2n} Y_n$$

$$\vdots$$

$$Z_n = l_{n1} Y_1 + l_{n2} Y_2 + \cdots + l_{nn} Y_n$$

where, l_{ij}'s are chosen such that the tramformation is orthogonal.

since, we get. $\underset{\sim}{Z} = L \underset{\sim}{Y}$, where L is an orthogonal mtx, so that, $|J| = 1$.

and $\sum_{i=1}^{n} Z_i^2 = Z'Z = Y'L'LY = Y'Y = \sum_{i=1}^{n} Y_i^2$

so the joint PDF of $Z_1, Z_2, \ldots, Z_n$ is

$$f_{Z_1,Z_2\ldots Z_n}(z_1, z_2, \ldots, z_n) = \left(\frac{1}{2\pi}\right)^{n/2} e^{-\frac{1}{2}\sum_{i=1}^{n} z_i^2}, \quad -\infty < z_i < \infty$$

i.e. $Z_1, Z_2, \ldots, Z_n$ are independently distributed as $N(0,1)$.

$$Z_1 = \frac{\sum_{i=1}^{n} Y_i \sqrt{\omega_i}}{\sqrt{\sum_{i=1}^{n} \omega_i}}$$

$$= \frac{\sum_{i=1}^{n} \omega_i (X_i - \mu)}{\sqrt{\sum_{i=1}^{n} \omega_i}} = \frac{\sum_{i=1}^{n} \omega_i X_i - \mu \sum_{i=1}^{n} \omega_i}{\sqrt{\sum_{i=1}^{n} \omega_i}}$$

$$= \frac{\bar{X}_\omega \sum_{i=1}^{n} \omega_i - \mu \sum_{i=1}^{n} \omega_i}{\sqrt{\sum_{i=1}^{n} \omega_i}}$$

$$= (\bar{X}_\omega - \mu) \sqrt{\sum_{i=1}^{n} \omega_i}$$

i.e. $$\bar{X}_\omega = \mu + \frac{1}{\sqrt{\sum_{i=1}^{n} \omega_i}} Z_1$$

since, $Z_1 \sim N(0,1)$

$$\bar{X}_\omega \sim N\left(\mu, \frac{1}{\sum_i \omega_i}\right), \quad \omega_i = \frac{1}{\sigma_i^2}$$

Also, $$\sum_{i=2}^{n} Z_i^2 = \sum_{i=1}^{n} Z_i^2 - Z_1^2$$

$$= \sum_{i=1}^{n} Y_i^2 - (\bar{X}_\omega - \mu)^2 \sum_{i=1}^{n} \omega_i$$

$$= \sum_{i=1}^{n} \omega_i (X_i - \mu)^2 - \sum_{i=1}^{n} \omega_i (\bar{X}_\omega - \mu)^2$$

$$= \sum_{i=1}^{n} \omega_i (X_i - \bar{X}_\omega)^2$$

$$= S_\omega^2$$

i.e. $S_\omega^2 = \sum_{i=2}^{n} Z_i^2$ = sum of squares of $(n-1)$ independent standard normal variables and hence follows χ^2_{n-1}.

Student's t-statistic & its distribution :

Let Z be a standard normal variate and χ^2_n be a chi-square variable with n degree of freedom such that Z is independent of χ^2_n then,

$$t = \frac{Z}{\sqrt{\chi^2/n}}$$, is called the student's statistic with n degree of freedom.

Distribution of t : To find the distn. of t, we start from the joint PDF of Z and χ^2 given by,

$$f_{Z\chi^2}(z,y) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \cdot \frac{1}{2^{n/2}\Gamma n/2} e^{-y/2} \cdot y^{\frac{n}{2}-1} I_x(-\infty,\infty) I_y(0,\infty)$$

Now let us take the tramformation,

$$(Z, \chi^2) \longrightarrow (R, \theta) \quad \text{such that,}$$

$$Z = r\sin\theta \qquad 0 < r < \infty$$

$$\sqrt{y} = r\cos\theta \qquad -\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

$$\Rightarrow y = r^2\cos^2\theta$$

Jacobian of the tramformation

$$J\left(\frac{z,y}{r,\theta}\right) = \begin{vmatrix} \sin\theta & r\cos\theta \\ 2r\cos^2\theta & -2r^2\sin\theta\cos\theta \end{vmatrix} = -2r^2\cos\theta$$

$$\therefore |J| = 2r^2\cos\theta \quad [\because r>0 \text{ and } \cos\theta > 0]$$

$\therefore$ Joint PDF of R and θ is

$$f_{R\theta}(r,\theta) = \frac{2r^2\cos\theta}{2^{\frac{n+1}{2}}\Gamma 1/2\, \Gamma n/2} \cdot e^{-\frac{1}{2}r^2} \cdot (r^2\cos^2\theta)^{\frac{n}{2}-1} I_r(0,\infty) I_\theta\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$$

$$= \frac{r^n \cdot e^{-r^2/2}}{\Gamma 1/2\, \Gamma n/2\; 2^{\frac{n-1}{2}}} (\cos\theta)^{n-1} I_r(0,\infty) I_\theta\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$$

Marginal PDF of θ,

$$f_\theta(\theta) = \int_0^\infty f_{R\theta}(r,\theta)\,dr = \frac{(\cos\theta)^{n-1}}{\overline{|1/2}\,\overline{|n/2}\;2^{\frac{n-1}{2}}}\int_0^\infty e^{-r^2/2}\,r^n\,dr$$

$$= \frac{(\cos\theta)^{n-1}\;2^{\frac{n-1}{2}}}{\overline{|1/2}\,\overline{|n/2}\;2^{\frac{n-1}{2}}}\int_0^\infty e^{-t}\,t^{\frac{n-1}{2}}\,dt \quad , \left[\frac{r^2}{2}=t \quad r\,dr = dt\right]$$

$$= \frac{(\cos\theta)^{n-1}}{\overline{|1/2}\,\overline{|n/2}}\cdot\overline{\left|\frac{n+1}{2}\right.}$$

$$= \frac{(\cos\theta)^{n-1}}{\beta\left(\frac{1}{2},\frac{n}{2}\right)}\, I_\theta\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$$

Now,

$$t = \frac{z}{\sqrt{y/n}} = \frac{z}{\sqrt{y}}\sqrt{n}$$

$$= \frac{r\sin\theta}{r\cos\theta}\sqrt{n}$$

$$= \sqrt{n}\tan\theta$$

$$\frac{dt}{d\theta} = \sqrt{n}\sec^2\theta \Rightarrow \frac{d\theta}{dt} = \frac{1}{\sqrt{n}\left(1+\frac{t^2}{n}\right)}$$

$$= \sqrt{n}\left(1+\frac{t^2}{n}\right)$$

$$\therefore (\cos^2\theta)^{\frac{n-1}{2}} = \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n-1}{2}}}$$

$$\therefore f_{t_n}(t) = \frac{1}{\beta\left(\frac{n}{2},\frac{1}{2}\right)}\cdot\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n-1}{2}}}\cdot\frac{1}{\sqrt{n}\left(1+\frac{t^2}{n}\right)}$$

$$= \frac{1}{\sqrt{n}\,\beta\left(\frac{n}{2},\frac{1}{2}\right)}\cdot\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}\; I_t(-\infty,\infty)$$

<u>Note:</u> – <u>limiting distribution of t for large n</u> $\longrightarrow$ [C.U. 2004]

$$f_{t_n}(t) = \frac{\overline{\left|\frac{n+1}{2}\right.}}{\sqrt{n}\,\overline{|n/2}\,\overline{|1/2}\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}.$$

$$= \frac{1}{\sqrt{n\pi}}\left(\frac{n}{2}\right)^{1/2}\cdot\frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \qquad \left[\because \frac{\overline{|n+k}}{\overline{|n}} \simeq n^k \text{ for large } n.\right]$$

$$= \frac{1}{\sqrt{2\pi}\left(1+t^2/n\right)^{n/2}\left(1+t^2/n\right)^{1/2}}$$

Now,

$$\lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{1/2} = 1$$

$$\& \lim_{n\to\infty}\left(1+\frac{t^2}{n}\right)^{n/2} = \lim_{n\to\infty}\left[1+\frac{t^2/2}{n/2}\right]^{n/2} = e^{t^2/2}$$

$$\therefore \text{for large } n, \quad f_{t_n}(t) \simeq \frac{1}{\sqrt{2\pi}}e^{-t^2/2} = \varphi(t).$$

Moments of t-distribution :

$$E\left(t^{2r+1}\right) = 0$$

$$\mu = E(t) = 0$$

$$\mu_{2r} = E\left(t^{2r}\right) = \frac{1}{\sqrt{n}\,\beta\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{-\infty}^{\infty} \frac{t^{2r}\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{2}{\sqrt{n}\,\beta\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} \frac{t^{2r}\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{n}{\sqrt{n}\,\beta\left(\frac{n}{2}, \frac{1}{2}\right)} \int_{0}^{\infty} \frac{(nz)^{\frac{2r-1}{2}}\,dz}{(1+z)^{\frac{n+1}{2}}}$$

$\left[\text{where, } \frac{t^2}{n} = z, \Rightarrow \frac{1}{n} 2t\,dt = dz\right]$

$$= n^{r} \frac{\beta\left(r+\frac{1}{2}, \frac{n}{2} - r\right)}{\beta\left(\frac{1}{2}, \frac{n}{2}\right)}$$

$$\mu_2 = n \cdot \frac{\beta\left(\frac{3}{2}, \frac{n}{2} - 1\right)}{\beta\left(\frac{1}{2}, \frac{n}{2}\right)} = \frac{n}{n-2} = \frac{1}{1 - 2/n}$$

$$\mu_4 = n^2 \cdot \frac{\beta\left(\frac{5}{2}, \frac{n}{2} - 2\right)}{\beta\left(\frac{1}{2}, \frac{n}{2}\right)} = n^2 \frac{\Gamma\frac{5}{2}\,\Gamma\frac{n-4}{2}\,\Gamma\frac{n+1}{2}}{\Gamma\frac{n+1}{2}\,\Gamma\frac{1}{2}\,\Gamma\frac{n}{2}}$$

$$= \frac{n^2\,3}{(n-2)(n-4)}.$$

$$\therefore \beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$= \frac{3n^2/(n-2)(n-4)}{n^2/(n-2)^2}$$

$$= 3 \cdot \frac{1-2/n}{1-4/n}$$

Hence, the t-distribution is perfectly skewed and leptokurtic.

F-Statistic & It's distribution :—

If χ_1^2 and χ_2^2 are two independent Chi-square variable with n_1 and n_2 degree of freedom then,

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$$ is called the F-statistic.

Distribution of F statistic :— To find the distribution of F-statistic we start from the joint PDF of χ_1^2 and χ_2^2 is given by,

$$f_{\chi_1^2 \chi_2^2}(x_1^2, x_2^2) = \frac{(\chi_1^2)^{\frac{n_1}{2}-1} \cdot (\chi_2^2)^{\frac{n_2}{2}-1}}{2^{\frac{n_1+n_2}{2}} \Gamma\frac{n_1}{2} \Gamma\frac{n_2}{2}} \; ; 0 < \chi_i^2 < \infty \quad i = 1, 2.$$

Now, let us consider the following transformation,

$$(\chi_1^2, \chi_2^2) \longrightarrow (F, V) \ni$$

$$\left.\begin{aligned} F &= \frac{n_2 \chi_1^2}{n_1 \chi_2^2} \\ \text{and, } V &= \chi_2^2 \end{aligned}\right\} \Rightarrow \begin{aligned} \chi_1^2 &= \frac{n_1}{n_2} FV, \quad 0 < F < \infty \\ \chi_2^2 &= V, \quad 0 < V < \infty \end{aligned}$$

Jacobian of the transformation is

$$J = \begin{vmatrix} \frac{n_1}{n_2} V & \frac{n_1}{n_2} F \\ 0 & 1 \end{vmatrix} = \frac{n_1}{n_2} V$$

So, the joint PDF of F and V is

$$f_{FV}(F, v) = \frac{\left(\frac{n_1 F v}{n_2}\right)^{\frac{n_1}{n_2}-1} v^{\frac{n_2}{2}-1} \frac{n_1}{n_2} v}{2^{\frac{n_1+n_2}{2}} \Gamma n_1/2 \, \Gamma n_2/2} \times e^{-\frac{1}{2}\left(1+\frac{n_1}{n_2}F\right)v} \qquad \begin{aligned} 0 < F < \infty \\ 0 < v < \infty \end{aligned}$$

Hence, the marginal PDF of F is,

$$f_F(F) = \frac{\frac{n_1}{n_2}\left(\frac{n_1 F}{n_2}\right)^{\frac{n_1}{n_2}-1}}{2^{\frac{n_1+n_2}{2}} \Gamma\frac{n_1}{2} \Gamma\frac{n_2}{2}} \int_0^\infty e^{-\frac{1}{2}v\left(1+\frac{n_1}{n_2}F\right)} v^{\frac{n_1+n_2}{2}-1} dv$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2} F^{\frac{n_1}{2}-1}}{2^{\frac{n_1+n_2}{2}} \Gamma\frac{n_1}{2} \Gamma\frac{n_2}{2}} \cdot \frac{\Gamma\frac{n_1+n_2}{2} \cdot 2^{\frac{n_1+n_2}{2}}}{\left(1+\frac{n_1 F}{n_2}\right)^{\frac{n_1+n_2}{2}}}, \quad 0 < F < \infty$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2} \cdot F^{\frac{n_1}{n_2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1 F}{n_2}\right)^{\frac{n_1+n_2}{2}}}, \quad 0 < F < \infty .$$

Note :-

$$X = \frac{\frac{n_1 F}{n_2}}{1 + \frac{n_1 F}{n_2}} \sim \beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$$

$$\Rightarrow n_2 X + n_1 F X = n_1 F$$

$$\Rightarrow F = \frac{n_2 X}{n_1(1-X)}.$$

$$\therefore |J| = \left|\frac{dF}{dx}\right| = \frac{n_2}{n_1}\left[\frac{1}{1-x} + \frac{x}{(1-x)^2}\right]$$

$\therefore$ PDF of X is,

$$f_X(x) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}-1} \cdot \left(\frac{x}{1-x}\right)^{\frac{n_1}{2}-1} \left(\frac{n_2}{n_1}\right)\left[\frac{1}{1-x} + \frac{x}{(1-x)^2}\right]}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left\{1 + \frac{n_1}{n_2} \cdot \frac{n_2}{n_1} \cdot \frac{x}{1-x}\right\}^{\frac{n_1+n_2}{2}}}$$

$$= \frac{x^{\frac{n_1}{2}-1}(1-x)^{\frac{n_2}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}$$

$$\therefore X \sim \beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right) ; 0 < x < 1$$

PDF of F can be also written as

$$f_F(F) = \left(\frac{n_1}{n_2}\right)\left(\frac{n_1 F}{n_2}\right)^{\frac{n_1}{2}-1} \Big/ \beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1 F}{n_2}\right)^{\frac{n_1+n_2}{2}}, \quad 0 < F < \infty.$$

X ——— X

BIVARIATE NORMAL DISTRIBUTION

Two jointly distributed continuous variables X & Y are distributed as bivariate normal variables with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ if their joint PDF is given by,

$$f_{XY}(x,y) = c\exp\left[-\frac{Q(x,y)}{2}\right]$$

where, $c = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$,

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Marginal Distribution of X and Y :—

The marginal PDF of X is given by, $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$.

Before integrating $f_{XY}(x,y)$ w.r.t. y, let us write $Q(x,y)$ in $f_{XY}(x,y)$ as follows:

$$Q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2} - \rho\,\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]$$

$$= \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \frac{1}{1-\rho^2}\left(\frac{y-\mu_2}{\sigma_2} - \rho\,\frac{x-\mu_1}{\sigma_1}\right)^2$$

so, $f(x) = c\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}Q(x,y)\right]dy$

$$= c\,e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left\{y-\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}^2} dy$$

$$= c.e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}.\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_2$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} I_x(-\infty,\infty).$$

i.e. marginal distribution of X is univariate normal (μ_1, σ_1^2).
Similarly it can be shown that, $Y \sim N(\mu_2, \sigma_2^2)$.

Conditional distribution of Y for given $X = x$:

$$f_{Y/X}(y/x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)}}{\frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}$$

$$= \frac{e^{-\frac{1}{2(1-\rho^2)\sigma_2^2}\left\{y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}^2}}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}}$$

i.e. for given $X = x$,

$$Y/X=x \sim N\left(\mu_2,\ \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$$

Note that,

$$E\left[Y/X=x\right] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),$$

which gives the regression of Y on X.

For two jointly distributed bivariate normal variables X and Y, the regression of Y on $X = x$ is linear.

Conditional variance of Y on X is,

$$\sigma_{YX}^2 = (1-\rho^2)\sigma_2^2 \quad \text{and} \quad \sigma_{YX}^2 \geq 0$$

$$\Rightarrow \rho^2 \leq 1, \text{ i.e. } |\rho| \leq 1.$$

Correlation coefficient of X and Y :

$E(X) = \mu_1$, $E(Y) = \mu_2$, $V(X) = \sigma_1^2$, $V(X) = \sigma_2^2$

$$\rho_{XY} = \frac{E(X-\mu_1)(Y-\mu_2)}{\sigma_1\sigma_2} = E\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]dx\,dy$$

$$U = \frac{X-\mu_1}{\sigma_1} \Rightarrow dx = \sigma_1\,du$$

$$V = \frac{Y-\mu_2}{\sigma_2} \Rightarrow dy = \sigma_2\,dv$$

$$\therefore \rho_{XY} = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv\exp\left[-\frac{1}{2(1-\rho^2)}\left\{u^2 - 2\rho uv + v^2\right\}\right]du\,dv$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv\exp\left[-\frac{u^2}{2}\right]\exp\left[-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right]du\,dv$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \left\{ \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} v \exp\left[-\frac{1}{2(1-\rho^2)}(v-\rho u)^2\right] dv \right\} e^{-u^2/2} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \rho\, u^2 e^{-u^2/2} du \qquad \left[\because E(Y) = \int_{-\infty}^{\infty} v \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2} dv, \right.$$

$$= \rho . \qquad \left. \text{where } V \sim N(\rho\mu, 1-\rho^2) \right]$$

i.e. $\rho_{XY} = \rho$

$\rho = 0$

$$\Rightarrow f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2}$$

i.e. X and Y are independent.

Equi Probability Contour of Bivariate Normal distribution: —

Let, $S_K = \{(x,y) : f_{XY}(x,y) = k\}$

Now, $f_{XY}(x,y) = k$

$$\Rightarrow \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \rho_2\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right] = 2\pi k\sigma_1\sigma_2\sqrt{1-\rho^2}$$

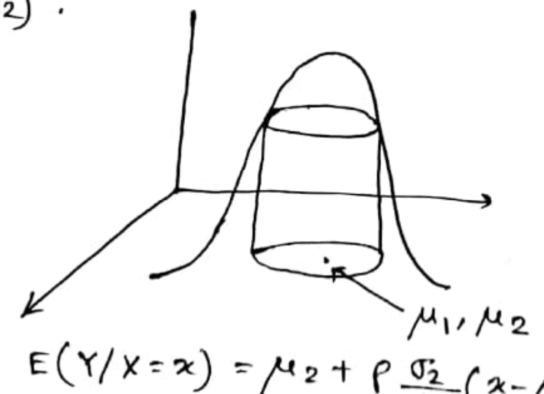
$$\therefore \left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 = -2(1-\rho^2)\ln\left(2\pi k\sigma_1\sigma_2\sqrt{1-\rho^2}\right)$$

$$= k', \text{ say}$$

$$\text{so, } S_K = \left\{(x,y) : \left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 = k'\right\}$$

where, $k' = 2(1-\rho^2)\ln\left(2\pi k\sigma_1\sigma_2\sqrt{1-\rho^2}\right)$.

i.e. the set of equiprobability contour, is an ellipse with centre at (μ_1, μ_2).



if $\rho = 1$, $E(Y/X=x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$

$$P\left[Y = \mu_2 + \frac{\sigma_2}{\sigma_1}(X-\mu_1)\right] = 1$$

$$\Rightarrow P\left[\frac{Y-\mu_2}{\sigma_2} = \frac{X-\mu_1}{\sigma_1}\right] = 1 .$$

☆ **Problem:—**

If $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. find the distribution of

$$U = \frac{X-\mu_1}{\sigma_1} \quad \& \quad V = \frac{1}{\sqrt{1-\rho^2}}\left\{\left(\frac{Y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{X-\mu_1}{\sigma_1}\right)\right\}$$

Ans:— Let us start from the joint PDF of X & Y given by,

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

Let us take the tramformation,

$$(X, Y) \longrightarrow (U, V) \ni$$

$$U = \frac{X-\mu_1}{\sigma_1} \quad \& \quad V = \frac{1}{\sqrt{1-\rho^2}}\left[\left(\frac{Y-\mu_2}{\sigma_2}\right) - \rho\left(\frac{X-\mu_1}{\sigma_1}\right)\right]$$

$$U^2 + V^2 = \frac{1}{(1-\rho^2)}\left[\left(\frac{X-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{X-\mu_1}{\sigma_1}\right)\left(\frac{Y-\mu_2}{\sigma_2}\right) + \left(\frac{Y-\mu_2}{\sigma_2}\right)^2\right]$$

$$X = \mu_1 + \sigma_1 U$$

$$Y = \mu_2 + \left[\rho u + \left(\sqrt{1-\rho^2}\right) V\right]\sigma_2$$

Jacobian of the tramformation is,

$$J = \begin{vmatrix} \sigma_1 & 0 \\ 0 & \sigma_2\sqrt{1-\rho^2} \end{vmatrix} = \sigma_1\sigma_2\sqrt{1-\rho^2}$$

$\therefore$ Joint PDF of U and V is

$$f_{UV}(u,v) = \frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)}, \quad -\infty < u, v < \infty$$

$$\therefore U = \frac{X-\mu_1}{\sigma_1} \sim N(0,1)$$

$$\therefore V = \frac{1}{\sqrt{1-\rho^2}}\left[\frac{Y-\mu_2}{\sigma_2} - \rho\frac{X-\mu_1}{\sigma_1}\right] \sim N(0,1)$$

$\therefore$ U and V are independent.

Sampling from a Bivariate Normal Population

Let $\{(X_i, Y_i) : i = 1(1)n\}$ be a paired sample of size n drawn from a bivariate normal $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then the joint PDF of $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$ is

$$f_{\underset{\sim}{X}\underset{\sim}{Y}}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$$

$$= \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)^n \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\sum_{i=1}^n \left(\frac{x_i-\mu_1}{\sigma_1}\right)^2 + \sum_{i=1}^n \left(\frac{y_i-\mu_2}{\sigma_2}\right)^2 - 2\rho\sum_{i=1}^n \left(\frac{x_i-\mu_1}{\sigma_1}\right)\left(\frac{y_i-\mu_2}{\sigma_2}\right)\right\}\right]$$

$-\infty < x_i, y_i < \infty$ for $i = 1(1)n$.

let us take the tramformation

$$(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n) \longrightarrow (\xi_1, \xi_2, \ldots \xi_n, \eta_1, \eta_2, \ldots, \eta_n)$$

$$\ni \quad \xi_i = \frac{X_i - \mu_1}{\sigma_1} \quad \& \quad \eta_i = \frac{1}{\sqrt{1-\rho^2}}\left\{\frac{Y_i - \mu_2}{\sigma_2} - \rho\frac{X_i - \mu_1}{\sigma_1}\right\} \forall\, i = 1(1)n$$

then the joint PDF of $(\xi_1, \xi_2, \ldots, \xi_n, \eta_1, \ldots, \eta_n)$ is

$$f_{\underset{\sim}{\xi}\underset{\sim}{\eta}}(\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n) = \left(\frac{1}{2\pi}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n(\xi_i^2 + \eta_i^2)}, \quad -\infty < \xi_i, \eta_i < \infty$$

$$= g(\xi_1, \xi_2, \ldots, \xi_n)\, h(\eta_1, \ldots \eta_n / \xi_1, \ldots \xi_n) \quad \forall\, i = 1(1)n.$$

where, $g(\xi_1, \ldots, \xi_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^n \xi_i^2\right], \quad -\infty < \xi_i < \infty$

and $h(\eta_1, \ldots, \eta_n / \xi_1, \ldots, \xi_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^n \eta_i^2\right], -\infty < \eta_i < \infty$

now, let us take the joint marginal distribution of $\xi_1, \ldots, \xi_n$. and consider the tramformation $(\xi_1, \xi_2, \ldots, \xi_n) \to (U_1, \ldots, U_n) \ni$

$$U_1 = \frac{1}{\sqrt{n}}(\xi_1 + \cdots + \xi_n)$$

$$\vdots$$

$$U_j = c_{j1}\xi_1 + c_{j2}\xi_2 + \cdots + c_{jn}\xi_n \quad \forall\, j = 2(1)n$$

where, c_{jk}'s are chosen $\ni$ the whole tramformation is orthogonal. Then the joint PDF of $U_1, U_2, \ldots, U_n$ is

$$g^*(u_1, u_2, \ldots, u_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^n u_i^2\right], \quad -\infty < u_i < \infty \quad \text{—— (1)}$$

Next, we consider the joint conditional PDF of $(\eta_1, \eta_2, \dots, \eta_n)$ for given $\xi_1, \xi_2, \dots, \xi_n$ and consider the transformation

$$(\eta_1, \eta_2, \dots, \eta_n / \xi_1, \xi_2, \dots, \xi_n) \longrightarrow (v_1, \dots, v_n / \xi_1, \dots, \xi_n)$$

$$v_1 = \frac{1}{\sqrt{n}} (\eta_1, \dots, \eta_n)$$

$$v_2 = \frac{1}{\sqrt{\sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}} \left[(\xi_1 - \bar{\xi})\eta_1 + (\xi_2 - \bar{\xi})\eta_2 + \dots + (\xi_n - \bar{\xi})\eta_n \right]$$

$$v_j = d_{j1}\eta_1 + \dots + d_{jn}\eta_n \;,\; j = 3(1)n.$$

d_{jk}'s are chosen $\ni$ the whole transformation became orthogonal. then the joint conditional PDF of $v_1, \dots, v_n$ gives $\xi_1, \dots, \xi_n$ which is same as the joint conditional PDF of $v_1, \dots, v_n$ given $U_1, U_2, \dots, U_n$ is,

$$h^*(v_1, \dots, v_n / u_1, \dots, u_n) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} v_i^2\right] \quad — ②$$

$$-\infty < v_i < \infty$$

from ① and ②, the joint PDF of $U_1, \dots, U_n$ & $V_1, \dots, V_n$ is,

$$f^*(u_1, \dots, u_n, v_1, \dots, v_n) = g^*(u_1, \dots, u_n)\, h^*(v_1, \dots, v_n / u_1, \dots, u_n)$$

$$= \left(\frac{1}{2\pi}\right)^{n} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} (u_i^2 + v_i^2)\right]$$

$$= \frac{1}{2\pi} \exp\left[-\frac{1}{2}(u_1^2 + v_1^2)\right] \left(\frac{1}{2\pi}\right)^{n-1} \exp\left[-\frac{1}{2}\sum_{i=2}^{n} (u_i^2 + v_i^2)\right]$$

$$= f_1^*(u_1, v_1)\, f_2^*(u_2, \dots, u_n, v_2, \dots, v_n), \text{ say}$$

so, (u_1, v_1) is independent of $(u_2, \dots, u_n)$ & $(v_2, \dots, v_n)$.

and we also get, U_i & V_i are independent $N(0,1)$ variables.

Now,
$$U_1 = \sqrt{n}\, \bar{\xi} = \sqrt{n}\left(\frac{\bar{X} - \mu_1}{\sigma_1}\right)$$

$$V_1 = \sqrt{n}\,.\bar{\eta}$$

$$= \sqrt{\frac{n}{1-\rho^2}} \left(\frac{\bar{Y} - \mu_2}{\sigma_2} - \rho \frac{\bar{X} - \mu_1}{\sigma_1}\right)$$

i.e.
$$U_1^2 + V_1^2 = n\left(\frac{\bar{X} - \mu_1}{\sigma_1}\right)^2 + \frac{n}{1-\rho^2}\left\{\frac{\bar{Y} - \mu_1}{\sigma_2} - \rho \frac{\bar{X} - \mu_1}{\sigma_1}\right\}^2$$

$$= \frac{n}{1-\rho^2}\left[\left(\frac{\bar{X} - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\bar{X} - \mu_1}{\sigma_1}\right)\left(\frac{\bar{Y} - \mu_2}{\sigma_2}\right) + \left(\frac{\bar{Y} - \mu_2}{\sigma_2}\right)^2\right]$$

$$\bar{X} = \mu_1 + \frac{\sigma_1}{\sqrt{n}} u_1$$

$$\bar{Y} = \mu_2 + \left[\frac{\rho}{\sqrt{n}} u_1 + \sqrt{\frac{1-\rho^2}{n}} v_1\right]\sigma_2$$

$$\Rightarrow v_1 = \sqrt{\frac{n}{1-\rho^2}}\left[\frac{\bar{Y}-\mu_2}{\sigma_2} - \rho\frac{u_1}{\sqrt{n}}\right]$$

Jacobian of the transformation is,

$$\frac{1}{|J|} = \begin{vmatrix} \sigma_1/\sqrt{n} & 0 \\ 0 & \sigma_2\sqrt{\frac{1-\rho^2}{n}} \end{vmatrix} = \frac{\sigma_1\sigma_2}{n}\sqrt{1-\rho^2}$$

Joint PDF of $\bar{X}, \bar{Y}$ is

$$Q_{\bar{X}\bar{Y}}(\bar{x},\bar{y}) = \frac{1}{2\pi} \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)\left(\frac{\bar{y}-\mu_2}{\sigma_2}\right) + \left(\frac{\bar{y}-\mu_2}{\sigma_2}\right)^2\right\}\right] \frac{n}{\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

$$= \frac{n}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{\bar{x}-\mu_1}{\sigma_1/\sqrt{n}}\right)^2 - 2\rho\left(\frac{\bar{x}-\mu_1}{\sigma_1/\sqrt{n}}\right)\left(\frac{\bar{y}-\mu_2}{\sigma_2/\sqrt{n}}\right) + \left(\frac{\bar{y}-\mu_2}{\sigma_2/\sqrt{n}}\right)^2\right\}\right]$$

$$-\infty < \bar{x}, \bar{y} < \infty .$$

$$s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Now, let us start from the joint PDF of $U_2, U_3, \ldots, U_n; V_2, \ldots, V_n$ is given by,

$$f^*(u_2, u_3, \ldots, u_n, v_2, v_3, \ldots, v_n) = \left(\frac{1}{2\pi}\right)^{n-1} e^{-\frac{1}{2}\sum_{i=2}^{n}(u_i^2 + v_i^2)}, \quad -\infty < u_i, v_i < \infty$$

Now, $T_1 = \sum_{i=2}^{n} U_i^2$

Then, T_1 being the sum of squares of $(n-1)$ independent standard normal variables follows χ^2_{n-1}.

$$T_1 = \sum_{i=2}^{n} U_i^2 = \sum_{i=1}^{n} U_i^2 - U_1^2 = \sum_{i=1}^{n} \xi_i^2 - n\bar{\xi}^2 = \sum_{i=1}^{n}(\xi_i - \bar{\xi})^2$$

$$= \frac{1}{\sigma_1^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{s_{xx}}{\sigma_1^2}.$$

$$\therefore \frac{s_{xx}}{\sigma_1^2} \sim \chi^2_{n-1}$$

$$T_2 = Y_2 = \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})\eta_i}{\sqrt{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}}$$

$$= \frac{\sum_{i=1}^{n}(\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\sqrt{\sum_{i=1}^{n}(\xi_i - \bar{\xi})^2}}$$

$$= \frac{1}{\sigma_1\sqrt{1-\rho^2}} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})\left[\frac{y_i - \bar{y}}{\sigma_2} - \rho\frac{x_i - \bar{x}}{\sigma_1}\right]}{\sqrt{\frac{S_{xx}}{\sigma_1^2}}}$$

$$= \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\left[S_{xy} - \rho\frac{\sigma_2}{\sigma_1}S_{xx}\right] \cdot \frac{1}{\sqrt{S_{xx}}/\sigma_1}$$

$$= \frac{1}{\sigma_2\sqrt{1-\rho^2}}\left(\frac{S_{xy} - \beta S_{xx}}{\sqrt{S_{xx}}}\right), \text{ where } \beta = \rho\frac{\sigma_2}{\sigma_1}$$

$\therefore Y_2 \sim N(0,1)$.

$$Y_2^2 = \frac{1}{\sigma_2^2(1-\rho^2)} \cdot \frac{(S_{xy} - \beta S_{xx})^2}{S_{xx}} \sim \chi_1^2$$

i.e. $T_2^2 = \frac{1}{\sigma_{2.1}^2}(S_{xy} - \beta S_{xx})^2 \cdot \frac{1}{S_{xx}} \sim \chi_1^2$

where $\sigma_{2.1}^2 = \sigma_2^2(1-\rho^2)$

let, $T_3 = \sum_{i=3}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 - Y_2^2$

$$= \sum_{i=1}^{n}\eta_i^2 - n\bar{\eta}^2 - Y_2^2$$

$$= \sum_{i=1}^{n}(\eta_i - \bar{\eta})^2 - Y_2^2$$

$$= \sum_{i=1}^{n}\frac{1}{1-\rho^2}\left[\frac{y_i - \bar{y}}{\sigma_2} - \rho\frac{x_i - \bar{x}}{\sigma_1}\right]^2 - Y_2^2$$

$$= \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n}\left\{(y_i - \bar{y}) - \rho\frac{\sigma_2}{\sigma_1}(x_i - \bar{x})\right\}^2 - Y_2^2$$

$$= \frac{1}{\sigma_{2.1}^2}\left[S_{yy} - 2\beta S_{xy} + \beta^2 S_{xx}\right] - Y_2^2$$

$$= \frac{1}{\sigma_{2.1}^2}\left[S_{yy} - \cancel{2\beta S_{xy}} + \cancel{\beta^2 S_{xx}} - \frac{S_{xy}^2}{S_{xx}} - \cancel{\beta^2 S_{xx}} + \cancel{2\beta S_{xy}}\right]$$

$$= \frac{1}{\sigma_{2.1}^2}\left[S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right]$$

$T_3 = \sum_{i=3}^{n} V_i^2$ = sum of squares of $(n-2)$ standard normal variables and has distribution χ^2_{n-2}.

$$= \frac{1}{\sigma_{2.1}^2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \sim \chi^2_{n-2}.$$

Now, we can find the joint distribution of S_{xx}, S_{yy}, S_{xy} starting from joint distribution of T_1, T_2, T_3 whose, PDF is given by,

$$f_{T_1T_2T_3}(T_1, T_2, T_3) = \frac{e^{-\frac{1}{2}[T_1 + T_2^2 + T_3]}(T_1)^{\frac{n-1}{2}-1}(T_3)^{\frac{n-2}{2}-1}}{\sqrt{2\pi}\, .\, 2^{\frac{n-1}{2}}\, 2^{\frac{n-2}{2}}\, \Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)},$$

$, T > 0$
$, -\infty < T_2 < \infty$
$, T_3 > 0$

Now, let us take the transformation,

$$T_1 = \frac{S_{xx}}{\sigma_1^2}$$

$$T_2 = \frac{(S_{xy} - \beta S_{xy})}{\sigma_{2.1}\sqrt{S_{xx}}}$$

$$T_3 = \frac{1}{\sigma_{2.1}^2}\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right)$$

Jacobian of the transformation is,

$$J = \left|J\left(\frac{T_1, T_2, T_3}{S_{xx}, S_{xy}, S_{yy}}\right)\right| = \begin{vmatrix} \frac{1}{\sigma_1^2} & 0 & 0 \\ * & \frac{1}{\sigma_{2.1}\sqrt{S_{xx}}} & 0 \\ * & * & \frac{1}{\sigma_{2.1}^2} \end{vmatrix}$$

$$= \frac{1}{\sigma_1^2\,\sigma_{2.1}^3\sqrt{S_{xx}}}$$

$\therefore T_1 + T_2^2 + T_3$

$$= \frac{S_{xx}}{\sigma_1^2} + \frac{(S_{xy} - \beta S_{xx})^2}{\sigma_{2.1}^2 S_{xx}} + \frac{\left(S_{yy} - \frac{S^2_{xy}}{S_{xx}}\right)}{\sigma_{2.1}^2}.$$

$$= \frac{S_{xx}}{\sigma_1^2} + \frac{1}{\sigma_{2.1}^2}\left[\cancel{\frac{S^2_{xy}}{S_{xx}}} - 2\beta\frac{\cancel{S_{xx}}S_{xy}}{\cancel{S_{xx}}} + \frac{\beta^2 S^2_{xx}}{S_{xx}} - \cancel{\frac{S^2_{xy}}{S_{xx}}} + S_{yy}\right]$$

$$= S_{xx}\left[\frac{1}{\sigma_1^2} + \frac{\beta^2}{\sigma_{2.1}^2}\right] + \frac{1}{\sigma_{2.1}^2}\left[S_{yy} - 2\beta S_{xy}\right]$$

$$= \frac{S_{xx}}{\sigma_1^2}\cdot\frac{1}{(1-\rho^2)} + \frac{1}{\sigma_2^2(1-\rho^2)}\left[S_{yy} - 2\beta S_{xy}\right]$$

$$= \frac{1}{1-\rho^2}\left[\frac{S_{xx}}{\sigma_1^2} - 2\rho\frac{S_{xy}}{\sigma_1\sigma_2} + \frac{S_{yy}}{\sigma_2^2}\right]$$

$\therefore$ Joint PDF of S_{xx}, S_{xy}, S_{yy} is,

$$f_{S_{xx}, S_{xy}, S_{yy}}(s_{xx}, s_{xy}, s_{yy})$$

$$= \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left[\frac{s_{xx}}{\sigma_1^2} - 2\rho\frac{s_{xy}}{\sigma_1\sigma_2} + \frac{s_{yy}}{\sigma_2^2}\right]\right]}{\sqrt{2\pi}\, 2^{\frac{2n-3}{2}} \Gamma\frac{n-1}{2}\Gamma\frac{n-2}{2}} \cdot \left(\frac{s_{xx}}{\sigma_1^2}\right)^{\frac{n-1}{2}-1} \left[\frac{s_{yy} - \frac{s_{xy}^2}{s_{xx}}}{\sigma_{2\cdot1}^2}\right]^{\frac{n-2}{2}-1} \cdot \frac{1}{\sigma_1^2\, \sigma_{2\cdot1}^3 \sqrt{s_{xx}}}$$

$$= c \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{s_{xx}}{\sigma_1^2} - 2\rho\frac{s_{xy}}{\sigma_1\sigma_2} + \frac{s_{yy}}{\sigma_2^2}\right]\right\}(s_{xx})^{\frac{n-2}{2}-1} \left[s_{yy} - \frac{s_{xy}^2}{s_{xx}}\right]^{\frac{n-2}{2}-1}$$

where,

$$C = \frac{1}{\sqrt{2\pi}\, 2^{\frac{2n-3}{2}} \Gamma\frac{n-1}{2}\Gamma\frac{n-2}{2}\, \sigma_1^{n-1}\, \sigma_{2\cdot1}^{n-1}}, \quad s_{xx} > 0,\ s_{yy} > 0,\ -\infty < s_{xy} < \infty$$

Distribution of Regression Coefficient :—

If $(X, Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$E[Y/X] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$
$$= \left(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1\right) + \rho\frac{\sigma_2}{\sigma_1}x$$
$$= \alpha + \beta x$$

$\beta = \rho\frac{\sigma_2}{\sigma_1}$ = the regression coefficient of Y on X, we can estimate this regression coefficient by sample regression coefficient.

$$b = \frac{s_{xy}}{s_{xx}}$$

$$T_2 = \frac{S_{xy} - \beta S_{xx}}{\sigma_{2\cdot1}\sqrt{S_{xx}}} = \frac{\sqrt{S_{xx}}\left(\frac{S_{xy}}{S_{xx}} - \beta\right)}{\sigma_{2\cdot1}} = \frac{\sqrt{S_{xx}}\,(b - \beta)}{\sigma_{2\cdot1}} \sim N(0,1)$$

$$b/_{S_{xx} = s_{xx}} = \beta + \frac{\sigma_{2\cdot1}}{\sqrt{s_{xx}}} T_2 \sim N\left(\beta_2, \frac{\sigma_2^2(1-\rho^2)}{s_{xx}}\right)$$

$$T_3 = \frac{1}{\sigma_{2\cdot1}^2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)$$

$$= \frac{1}{\sigma_{2\cdot1}^2}\left[S_{yy} - \frac{r^2 S_{xx} S_{yy}}{S_{xx}}\right]$$

$$= \frac{(1-r^2) S_{yy}}{\sigma_{2\cdot1}^2} \sim \chi^2_{n-2}$$

$$E\left[\frac{(1-r^2) S_{yy}}{\sigma_{2\cdot1}^2}\right] = n-2$$

$$\Rightarrow E\left[\frac{(1-r^2) S_{yy}}{n-2}\right] = \sigma_{2\cdot1}^2$$

$\frac{(1-r^2) S_{yy}}{n-2}$ is an unbiased estimate of $\sigma_{2\cdot1}^2$ and let be denoted by, $\hat{\sigma}_{2\cdot1}$.

So, $$\frac{\sqrt{S_{xx}}(b-\beta)}{\hat{\sigma}_{2\cdot1}} = \frac{\frac{\sqrt{S_{xx}}(b-\beta)}{\sigma_{2\cdot1}}}{\left(\hat{\sigma}_{2\cdot1}^2/\sigma_{2\cdot1}^2\right)^{1/2}} = \frac{\sqrt{S_{xx}}(b-\beta)/\sigma_{2\cdot1}}{\sqrt{\frac{(1-r^2)S_{yy}}{\sigma_{2\cdot1}^2}/n-2}}$$

$$= \frac{T_2}{\sqrt{T_3/n-2}}$$

$T_2 \sim N(0,1)$

$T_3 \sim \chi^2_{n-2}$, T_2, T_3 are independent.

$$\frac{\sqrt{S_{xx}}(b-\beta)}{\hat{\sigma}_{2\cdot1}} \sim t_{n-2}.$$

Distribution of correlation coefficient r :—

To find r, $r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$

so, to find the distribution of r from the joint distribution of S_{xx}, S_{xy}, S_{yy}.

Let us take the tranformation,

$$(S_{xx}, S_{xy}, S_{yy}) \longrightarrow (U, V, r)$$

$$S_{xx} = U$$
$$S_{yy} = V$$
$$S_{xy} = r\sqrt{UV}$$

Jacobian of the transformation is , $|J| = \sqrt{uv}$

So, the joint PDF of U, V, r is

$$f_{UVr}(u,v,r) = C\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{u}{\sigma_1^2} - \frac{2\rho r\sqrt{uv}}{\sigma_1\sigma_2} + \frac{v}{\sigma_2^2}\right]u^{\frac{n-2}{2}-1}(v - r^2 v)^{\frac{n-2}{2}-1}\sqrt{uv}$$

$$= C\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{u}{\sigma_1^2} + \frac{v}{\sigma_2^2} - 2\rho r\frac{\sqrt{uv}}{\sigma_1\sigma_2}\right\}\right](uv)^{\frac{n-1}{2}-1}(1-r^2)^{\frac{n-2}{2}-1}$$

$$f_r(r) = \int_0^\infty\int_0^\infty f_{UVr}(u,v,r)\,du\,dv \qquad -1<r<1,\ u,v>0$$

the distribution of r when, $\rho = 0$

$$f(r/\rho=0) = \frac{(1-r^2)^{\frac{n-2}{2}-1}\left(\Gamma\frac{n-1}{2}\right)^2(2\sigma_1^2)^{\frac{n-1}{2}}(2\sigma_2^2)^{\frac{n-1}{2}}}{2^{n-1}\sqrt{\pi}\,\Gamma\frac{n-1}{2}\,\Gamma\frac{n-2}{2}\,(\sigma_1,\sigma_2)^{n-1}}$$

$$= \frac{(1-r^2)^{\frac{n-2}{2}-1}}{\beta\left(\frac{1}{2},\frac{n-2}{2}\right)}, \quad -1<r<1$$

$$f_r(r) = \frac{(1-r^2)^{\frac{n-2}{2}-1}}{\beta\left(\frac{1}{2},\frac{n-2}{2}\right)}, \quad -1<r<1$$

Now, $\underbrace{t = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}}$ [To show]

Let us take the tramformation,

$$t = r\frac{\sqrt{n-2}}{\sqrt{1-r^2}}, \quad t \in \mathbb{R}$$

$$\Rightarrow r = \frac{1}{\sqrt{1+\frac{n-2}{t^2}}} = \frac{1}{\frac{\sqrt{n-2}}{t}\sqrt{\frac{t^2}{n-2}+1}}$$

Jacobian of tramformation,

$$J\left(\frac{r}{t}\right) = \frac{dr}{dt} = \frac{1}{(-\cancel{2})} \cdot \frac{1}{\left(1+\frac{n-2}{t^2}\right)^{3/2}} \cdot (n-2) \cdot t^{-3} \cdot (-\cancel{2})$$

$$= \frac{(n-2)}{\left(\frac{n-2}{t^2}\right)^{3/2}\left(\frac{t^2}{n-2}+1\right)^{3/2}} \cdot \frac{1}{t^3}$$

$$= \frac{1}{\sqrt{n-2}\left(\frac{t^2}{n-2}+1\right)^{3/2}}$$

$$\therefore f_t(t) = \frac{\left(1 - \frac{1}{1+\frac{n-2}{t^2}}\right)^{\frac{n-2}{2}-1}}{\beta\left(\frac{1}{2}, \frac{n-2}{2}\right)} \cdot \frac{1}{\sqrt{n-2}\left(\frac{t^2}{n-2}+1\right)^{3/2}}$$

$$= \frac{\left\{\frac{(\cancel{n-2})/\cancel{t^2}}{\frac{(\cancel{n-2})}{\cancel{t^2}}\left(\frac{t^2}{n-2}+1\right)}\right\}^{\frac{n-2}{2}-1}}{\beta\left(\frac{1}{2}, \frac{n-2}{2}\right)\sqrt{n-2}\left(\frac{t^2}{n-2}+1\right)^{3/2}}$$

$$= \frac{1}{\beta\left(\frac{1}{2}, \frac{n-2}{2}\right)\sqrt{n-2}\left(\frac{t^2}{n-2}+1\right)^{\frac{n-1}{2}}} \quad I_t(-\infty,\infty)$$

$$\therefore t \sim t_{n-2}$$

$$C = \frac{1}{2^{n-1}\sqrt{\pi}\,(\sigma_1\sigma_2)^{n-1}(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma\frac{n-1}{2}\,\Gamma\frac{n-2}{2}}$$

Now, we are to evaluate the PDF of r, so, the PDF of r is,

$$f(r/\rho) = C(1-r^2)^{\frac{n-4}{2}} \int_0^\infty\int_0^\infty (uv)^{\frac{n-3}{2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{u}{\sigma_1^2}+\frac{v}{\sigma_2^2} - 2\rho r\,\frac{\sqrt{uv}}{\sigma_1\sigma_2}\right\}\right] du\,dv$$

$$= C(1-r^2)^{\frac{n-4}{2}} \int_0^\infty\int_0^\infty (uv)^{\frac{n-3}{2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{u}{\sigma_1^2}+\frac{v}{\sigma_2^2}\right)\right] \exp\left[\frac{\rho r}{1-\rho^2}\cdot\frac{\sqrt{uv}}{\sigma_1\sigma_2}\right] du\,dv$$

$$= C(1-r^2)^{\frac{n-4}{2}} \int_0^\infty\int_0^\infty (uv)^{\frac{n-3}{2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{u}{\sigma_1^2}+\frac{v}{\sigma_2^2}\right)\right] \sum_{s=0}^\infty \left[\frac{\rho r}{\sigma_1\sigma_2(1-\rho^2)}\right]^s \frac{(uv)^{s/2}}{s!}\, du\,dv$$

$$\cancel{\boxed{= C(1-r^2)^{\frac{n-4}{2}} \sum_{s=0}^\infty \left[\frac{\rho r}{\sigma_1\sigma_2(1-\rho^2)}\right]^s \cdot \frac{1}{s!}\left\{\int_0^\infty e^{-\frac{u}{2(1-\rho^2)\sigma_1^2}} u^{\frac{n+s-1}{2}-1} du\right\} \left\{\int_0^\infty e^{-\frac{v}{2}}\right.}}$$

$$= C(1-r^2)^{\frac{n-4}{2}} \sum_{s=0}^\infty \left[\frac{\rho r}{\sigma_1\sigma_2(1-\rho^2)}\right]^s \cdot \frac{1}{s!} \int_0^\infty\int_0^\infty \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{u}{\sigma_1^2}+\frac{v}{\sigma_2^2}\right)\right] (uv)^{\frac{n+s-3}{2}} du\,dv$$

$$= C(1-r^2)^{\frac{n-4}{2}} \sum_{s=0}^\infty \left[\frac{\rho r}{\sigma_1\sigma_2(1-\rho^2)}\right]^s \cdot \frac{1}{s!} \left\{\int_0^\infty e^{-\frac{u}{2(1-\rho^2)\sigma_1^2}} u^{\frac{n+s-1}{2}-1} du\right\} \left\{\int_0^\infty e^{-\frac{v}{2(1-\rho^2)\sigma_2^2}} v^{\frac{n+s-1}{2}-1} dv\right\}$$

$$= \frac{(1-r^2)^{\frac{n-4}{2}} \sum_{s=0}^\infty \left[\frac{\rho r}{\sigma_1\sigma_2(1-\rho^2)}\right]^s \frac{1}{s!}\left(\Gamma\frac{n+s-1}{2}\right)^2 \left(2(1-\rho^2)\sigma_1\sigma_2\right)^{n+s-1}}{2^{n-1}\sqrt{\pi}\,(\sigma_1\sigma_2)^{n-1}(1-\rho^2)^{\frac{n-1}{2}}\,\Gamma\frac{n-1}{2}\,\Gamma\frac{n-2}{2}}$$

$$= \frac{(1-r^2)^{\frac{n-4}{2}}(1-\rho^2)^{\frac{n-1}{2}} \sum_{s=0}^\infty \frac{1}{s!}(2\rho r)^s \left(\Gamma\frac{n+s-1}{2}\right)^2}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}+\frac{1}{2}\right)\Gamma\frac{n-2}{2}}$$

$$= \frac{2^{n-3}(1-r^2)^{\frac{n-4}{2}}(1-\rho)^{\frac{n-1}{2}} \sum_{s=0}^\infty (2\rho r)^s \cdot \frac{1}{s!}\left(\Gamma\frac{n+s-1}{2}\right)^2}{\pi\,\Gamma(n-2)}$$

[Using legendre's Duplication formula]

$, |r| < 1$

Distribution of Regression Co-efficient :—

Let $x_1, x_2, \ldots, x_n$ be n fixed value taken by a non-stochastic variable x. Also let Y be a random variable such that for fixed x.

$$Y/x \sim N(\eta_x, \sigma^2) \text{ where, } \eta_x = E[Y/X] = \alpha + \beta x$$

η_x is the regression equation of y on x. Now, let the fitted regression line be,

$$Y = a + bX, \text{ where } b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad a = \bar{Y} - b\bar{x}$$

To find the distribution of a and b let us start from the joint PDF of $Y_1, Y_2, \ldots, Y_n$ given by,

$$f(y_1, y_2, \ldots, y_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right]; \quad -\infty < y_i < \infty \quad i = 1(1)n$$

Now, let $Z_i = Y_i - \alpha - \beta x_i \quad \forall i = 1(1)n$.

$$f_{\underset{\sim}{Z}}(z_1, \ldots, z_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} z_i^2\right], \quad -\infty < z_i < \infty \quad \forall i = 1(1)n.$$

Now, let us consider the transformation,

$$(Z_1, Z_2, \ldots, Z_n) \longrightarrow (U_1, U_2, \ldots, U_n) \ni$$

$$U_1 = \frac{1}{\sqrt{n}}(Z_1 + \cdots + Z_n)$$

$$U_2 = \frac{1}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\left[(x_1 - \bar{x})Z_1 + (x_2 - \bar{x})Z_2 + \cdots + (x_n - \bar{x})Z_n\right]$$

$$U_j = l_{j1}Z_1 + l_{j2}Z_2 + \cdots + l_{jn}Z_n, \quad j = 3(1)n.$$

l_{jk} are chosen such that the transformation are orthogonal.

Then, $\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n} Z_i^2 \quad \forall i = 1(1)n$

$$|J| = 1$$

$\therefore$ The joint PDF of $U_1, U_2, \ldots, U_n$ is

$$f_{\underset{\sim}{U}}(u_1, u_2, \ldots, u_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i^2\right], \quad -\infty < u_i < \infty \quad \forall i = 1(1)n$$

thus, $U_i \sim N(0, \sigma^2) \quad \forall i = 1(1)n$

U_i's are independent.

$$U_1 = \sqrt{n}\,\bar{Z} = \sqrt{n}\left(\bar{Y} - \alpha - \beta\bar{x}\right)$$

$$U_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Z_i}{\sqrt{\sum (x_i - \bar{x})^2}} = \frac{\sum (x_i - \bar{x})\left(Y_i - \alpha - \beta x_i\right)}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$= \frac{S_{xx}\, b - \beta S_{xx}}{\sqrt{S_{xx}}}$$

$$= \sqrt{S_{xx}}\,(b - \beta) \sim N(0, \sigma^2)$$

$$\Rightarrow \frac{(b-\beta)\sqrt{S_{xx}}}{\sigma} \sim N(0,1)$$

$$\therefore\ b = \beta + \frac{1}{\sqrt{S_{xx}}} U_2$$

$$\therefore\ b \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$$

$$U_1 = \sqrt{n}\left(\bar{Y} - \alpha - \beta\bar{x}\right) \sim N(0, \sigma^2)$$

$$\Rightarrow \bar{Y} = \alpha + \beta\bar{x} + \frac{1}{\sqrt{n}} U_1$$

$$a = \bar{Y} - b\bar{x}$$

$$= \alpha + \beta\bar{x} + \frac{1}{\sqrt{n}} U_1 - \left(\beta + \frac{U_2}{\sqrt{S_{xx}}}\right)\bar{x}$$

$$= a + \frac{U_1}{\sqrt{n}} - \frac{\bar{x}}{S_{xx}} U_2$$

i.e. 'a' being a linear combination of two independent normal variables also follows normal distribution.

$$E(a) = \alpha$$

$$V(a) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

i.e. $a \sim N\left(\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$

if σ is unknown, —

$$\sum_{i=3}^{n} U_i^2 = \sum_{i=1}^{n} U_i^2 - U_1^2 - U_2^2$$

$$= \sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2 - \frac{\left\{\sum_{i=1}^{n}(x_i - \bar{x})Z_i\right\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sum_{i=1}^{n}(Z_i - \bar{Z})^2 - \frac{\left\{\sum_{i=1}^{n}(x_i - \bar{x})Z_i\right\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sum_{i=1}^{n}\left[(Y_i - \bar{Y}) - \beta(x_i - \bar{x})\right]^2 - \frac{\left\{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \alpha - \beta x_i)\right\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= S_{YY} + \beta^2 S_{xx} - 2\beta S_{xy} - \frac{(S_{xy} - \beta S_{xx})^2}{S_{xx}}$$

$$= S_{YY} - \frac{S_{xy}^2}{S_{xx}}$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 - \frac{S_{xY}}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) Y_i$$

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 - b \sum_{i=1}^{n} (x_i - \bar{x}) Y_i$$

$$= \sum_{i=1}^{n} (Y_i - a - bx_i)^2$$

$$= (n-2) s_{Y.x}^2 \quad \text{, say}$$

So,
$$\frac{(n-2) s_{Y.x}^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{U_i}{\sigma}\right)^2 \sim \chi_{n-2}^2$$

[independent of U_1, U_2 i.e. a, b]

$$\therefore E\left[\frac{(n-2) s_{Y.x}^2}{\sigma^2}\right] = n-2$$

$$\therefore E\left[s_{Y.x}^2\right] = \sigma^2$$

$$\frac{(b-\beta)\sqrt{S_{xx}}}{s_{Y.x}} = \frac{(b-\beta)\sqrt{S_{xx}}/\sigma}{\sqrt{\frac{(n-2)s_{Y.x}^2}{\sigma^2}/n-2}} = \frac{N}{\sqrt{\chi_{n-2}^2/n-2}} \sim t_{n-2},$$

where, $N \sim N(0,1)$.

⇨ The distribution of regression coefficient b when X is a stochastic variable.

What happens when X is a stochastic variable such that $(X,Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

Since, $(X,Y) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

$$Y/X=x \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$$

So, from the earlier result

$$b/X=x \sim N\left(\beta,\ \frac{\sigma_2^2(1-\rho^2)}{S_{xx}}\right)$$

or, $$b/S_{xx} \sim N\left(\beta,\ \frac{\sigma_2^2(1-\rho^2)}{S_{xx}}\right)$$

To find the unconditional distribution of b, we note that

$$\frac{S_{xx}}{\sigma_1^2} \sim \chi^2_{n-1}$$

So, the unconditional PDF of b is

$$f(b) = \int_0^\infty f_1(b/S_{xx})\, f_2(S_{xx})\, dS_{xx}$$

$$= \int_0^\infty \frac{\sqrt{S_{xx}}}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{S_{xx}}{2\sigma_2^2(1-\rho^2)}(b-\beta)^2\right] \cdot \frac{\exp\left[-\frac{S_{xx}}{2\sigma_1^2}\right]}{2^{\frac{n-1}{2}}\Gamma\frac{n-1}{2}} \cdot \frac{1}{\sigma_1^2}\left(\frac{S_{xx}}{\sigma_1^2}\right)^{\frac{n-1}{2}} dS_{xx}$$

$$= \frac{1}{\sigma_1^{n-1}\, 2^{n/2}\sqrt{\pi}\,\sigma_2\sqrt{1-\rho^2}\,\Gamma\frac{n-1}{2}} \int_0^\infty \exp\left[-\frac{1}{2\sigma_1^2}\left\{1+\frac{\sigma_1^2(b-\beta)^2}{\sigma_2^2(1-\rho^2)}\right\}S_{xx}\right] (S_{xx})^{n/2-1}\, dS_{xx}$$

$$= \frac{\Gamma\frac{n}{2}\,(2\sigma_1^2)^{n/2}}{2^{n/2}\sqrt{\pi}\,\sigma_2\sqrt{1-\rho^2}\,\Gamma\frac{n-1}{2}\,\sigma_1^{n-1}\left[1+\frac{\sigma_1^2(b-\beta)^2}{\sigma_2^2(1-\rho^2)}\right]^{n/2}}$$

$$= \frac{\sigma_1}{\beta\left(\frac{n-1}{2}, \frac{1}{2}\right)\sigma_2\sqrt{1-\rho^2}} \cdot \frac{1}{\left[1+\frac{\sigma_1^2(b-\beta)^2}{\sigma_2^2(1-\rho^2)}\right]^{n/2}}$$

let, $t = \dfrac{\sigma_1 (b-\beta)}{\sigma_2 (\sqrt{1-\rho^2})} \sqrt{n-1}$.

$\therefore \ |J| = \dfrac{\sigma_2 \sqrt{1-\rho^2}}{\sigma_1 \sqrt{n-1}}$

$$\therefore f(t) = \frac{1}{\sqrt{n-1}\, \beta\left(\frac{n-1}{2}, \frac{1}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{n-1}\right)^{n/2}}$$

i.e. if X is stochastic then $t = \dfrac{\sigma_1 (b-\beta)\sqrt{n-1}}{\sigma_2 \sqrt{1-\rho^2}} \sim t_{n-1}$.

PAIRED 't' STATISTIC :— Definition :—

Let $(X_1, Y_1), (X_2, Y_2), \ldots\ldots, (X_n, Y_n)$ are random samples drawn from bivariate normal $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$.

Let us define , $Z_i = X_i - Y_i$

$$\therefore \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$$

$\bar{Z} \sim N\left(\mu_Z, \dfrac{\sigma_Z^2}{n}\right)$, where $\mu_Z = \mu_x - \mu_y$ and

$$\sigma_Z^2 = V(X_i - Y_i)$$
$$= \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$$

Let us define, $s_Z^2 = \dfrac{1}{n}\sum_{i=1}^{n} (Z_i - \bar{Z})^2$

$$= \frac{1}{n}\sum_{i=1}^{n} \left\{ (X_i - \bar{X}) - (Y_i - \bar{Y}) \right\}^2$$

$$= s_x^2 + s_y^2 - 2 r s_x s_y$$

Define, $t = \dfrac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z} \sim t_{n-1}$

then $t = \dfrac{\sqrt{n}(\bar{Z} - \mu_Z)}{s_Z}$ is called paired 't' statistic.

BIVARIATE NORMAL DISTRIBUTION :—

Let X and Y be continuous r.v. jointly distributed over $\mathbb{R}^2$ having joint PDF 'f', where

$$f(x,y) = \text{Constant}.\, e^{-\frac{1}{2}Q(x,y)}, \quad -\infty < x, y < \infty$$

$Q(x,y) = ax^2 - 2hxy + by^2$; $\quad$ a, b, h are suitable constants.

[$Q(x,y)$ = Positive definite quadratic form]

Let, $f(x,y) = K.e^{-\frac{1}{2}Q(x,y)}$

Determination of constant :—

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

$$\Rightarrow K\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(ax^2-2hxy+by^2)}\,dx\,dy = 1.$$

$$\Rightarrow K\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{-\frac{a}{2}\left(x-\frac{hy}{a}\right)^2}dx\right\} e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}dy = 1$$

$$\Rightarrow K.\frac{\sqrt{2\pi}}{\sqrt{a}}\int_{-\infty}^{\infty} e^{-\frac{ab-h^2}{a}\left(\frac{y^2}{2}\right)}dy = 1$$

$$\Rightarrow K.\sqrt{\frac{2\pi}{a}}\cdot\sqrt{2\pi}\cdot\sqrt{\frac{a}{ab-h^2}} = 1$$

$$\Rightarrow K = \frac{\sqrt{ab-h^2}}{2\pi}$$

$$E(X) = K\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x e^{-\frac{1}{2}Q(x,y)}\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\frac{x e^{-\frac{a}{2}\left(x-\frac{hy}{a}\right)^2}}{\sqrt{\frac{2\pi}{a}}}dx\right\}\frac{e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}dy$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}\cdot\frac{hy}{a}.e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}dy$$

$$= \frac{h}{a}\times 0$$

$$= 0$$

Similarly, $E(Y) = 0$; $Q(x,y)$ = homogeneous 2nd degree in x & y.

Scanned by CamScanner

Note:— $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

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

$$= \mu^2 + \sigma^2 \cdot \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} z^2 e^{-z^2/2}\, dz$$

$$= \mu^2 + \sigma^2 \cdot \left(\frac{\cancel{2}}{\sqrt{2\cancel{\pi}}} \cdot \cancel{\sqrt{2}} \cdot \frac{1}{\cancel{2}}\cancel{\sqrt{\pi}}\right)$$

$$= \mu^2 + \sigma^2 .$$

Now, $\sigma_x^2 = V(X) = k\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}Q(x,y)}\, dx\, dy$

$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x^2 \cdot \frac{e^{-\frac{a}{2}\left(x-\frac{hy}{a}\right)^2}}{\sqrt{\frac{2\pi}{a}}}\cdot dx\right\}\frac{e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}\, dy$$

$$= \int_{-\infty}^{\infty}\left(\frac{h^2y^2}{a^2}+\frac{1}{a}\right)\frac{e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}\, dy$$

$$= \frac{1}{a} + \frac{h^2}{a^{\cancel{2}}}\cdot\frac{\cancel{a}}{ab-h^2}$$

$$= \frac{b}{ab-h^2} .$$

Similarly, $\sigma_y^2 = V(Y) = \frac{a}{ab-h^2}$.

$Cov(X,Y) = \rho\sigma_x\sigma_y$

$$= E(XY)$$

$$= k\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, e^{-\frac{1}{2}Q(x,y)}\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} x \cdot \frac{e^{-\frac{a}{2}\left(x-\frac{hy}{a}\right)^2}}{\sqrt{\frac{2\pi}{a}}}\, dx\right\} y \cdot \frac{e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}\, dy$$

$$= \int_{-\infty}^{\infty}\frac{hy}{a}\cdot y \cdot \frac{e^{-\frac{ab-h^2}{a}\cdot\frac{y^2}{2}}}{\sqrt{2\pi}\sqrt{\frac{a}{ab-h^2}}}\, dy$$

$$= \frac{h}{\cancel{a}}\cdot\frac{\cancel{a}}{ab-h^2}$$

$$= \frac{h}{ab-h^2} .$$

Note that, $(1-\rho^2)\sigma_x^2\sigma_y^2$

$$= \sigma_x^2\sigma_y^2 - \rho^2\sigma_x^2\sigma_y^2$$

$$= \frac{ab}{(ab-h^2)^2} - \frac{h^2}{(ab-h^2)^2}$$

$$= \frac{1}{(ab-h^2)}$$

$$b = (ab-h^2)\sigma_x^2$$

$$= \frac{1}{(1-\rho^2)\sigma_y^2}$$

$$a = (ab-h^2)\sigma_y^2$$

$$= \frac{1}{(1-\rho^2)\sigma_x^2}$$

$$h = \rho\sigma_x\sigma_y(ab-h^2)$$

$$= \frac{\rho}{(1-\rho^2)\sigma_x\sigma_y}$$

$$k = \frac{\sqrt{ab-h^2}}{2\pi} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$\therefore f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_x}\right)^2+\left(\frac{y}{\sigma_y}\right)^2-2\rho\left(\frac{x}{\sigma_x}\right)\left(\frac{y}{\sigma_y}\right)\right\}}$$

$$, -\infty < x, y < \infty$$

Define,

$X' = X - \mu_x$

$Y' = Y - \mu_y$

Joint PDF of (X', Y') will be

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right\}}$$

$$(x,y)\in\mathbb{R}^2; (\mu_x,\mu_y)\in\mathbb{R}^2$$

$$\sigma_x>0, \sigma_y>0, |\rho|<1$$

PDF of bivariate normal distribution with means μ_x and μ_y and variances σ_x^2 and σ_y^2 and correlation coefficient ρ, we denote by, $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

~~Alternative Representation :— (as a special case of multivariate normal distribution)~~

~~$\underset{\sim}{\mu}$ = mean vector = $\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$~~

~~dispersion matrix :—~~

~~$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$ {var-cov matrix}~~

~~PDF of $\underset{\sim}{X} = (X_1, \ldots, X_p)' \sim$ p-variate normal with mean vector~~

Determination of Constants (alternative way) :—

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(ax^2 - 2hxy + by^2)}\,dx\,dy = \frac{2\pi}{\sqrt{ab-h^2}}$$

Partial derivative w.r.t. a,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -\frac{x^2}{2} e^{-\frac{1}{2}(ax^2 - 2hxy + by^2)}\,dx\,dy = \frac{2\pi b}{-2(ab-h^2)^{3/2}}$$

$$= \frac{2\pi}{\sqrt{ab-h^2}}\left(\frac{-b}{2(ab-h^2)}\right)$$

$$\therefore \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x,y)\,dx\,dy = \frac{b}{ab-h^2}$$

Partial derivative w.r.t. b,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -\frac{y^2}{2} e^{-\frac{1}{2}(ax^2 - 2hxy + by^2)}\,dx\,dy = \frac{2\pi a}{-2(ab-h^2)^{3/2}}$$

$$= \frac{2\pi}{\sqrt{ab-h^2}}\left(\frac{-a}{2(ab-h^2)}\right)$$

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

Partial derivative w.r.t. h,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\cancel{2}xy}{\cancel{2}} e^{-\frac{1}{2}(ax^2 - 2hxy + by^2)}\,dx\,dy = \frac{2\pi \cdot \cancel{2}h}{\cancel{2}\sqrt{ab-h^2}\,(ab-h^2)}$$

$$\Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy = \frac{h}{ab-h^2}$$

$$\therefore \sigma_x^2 = \frac{b}{ab-h^2}, \quad \sigma_y^2 = \frac{a}{ab-h^2}$$

$$\mathrm{Cov}(X,Y) = \rho\sigma_x\sigma_y = \frac{h}{ab-h^2}$$

Moment Generating function :

$$(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

$$M_{XY}(t,s) = E\left(e^{tX+sY}\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp\left\{tx+sy-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\,dx\,dy$$

$(x,y) \longrightarrow (u,\theta)$ $\ni$

$$u = \frac{x-\mu_x}{\sigma_x}, \quad v = \frac{y-\mu_y}{\sigma_y}$$

$$J\left(\frac{x,y}{u,\theta}\right) = \begin{vmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{vmatrix} = \sigma_x\sigma_y$$

$$M_{XY}(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{t(\mu_x+u\sigma_x)+s(\mu_y+v\sigma_y)} \cdot e^{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)}}{2\pi\sqrt{1-\rho^2}}\,du\,dv$$

$$= e^{t\mu_x+s\mu_y}\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\frac{e^{ut\sigma_x-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}}{\sqrt{2\pi}\sqrt{1-\rho^2}}\,du\right\}\frac{e^{sv\sigma_y-\frac{v^2}{2}}}{\sqrt{2\pi}}\,dv$$

$$= e^{t\mu_x+s\mu_y}\int_{-\infty}^{\infty} e^{\rho vt\sigma_x+(1-\rho^2)\frac{t^2\sigma_x^2}{2}} \cdot e^{sv\sigma_y-\frac{v^2}{2}}\,dv$$

$\left[\because M_X(t) = e^{\mu t+\frac{\sigma^2 t^2}{2}},\right.$ when $X \sim N(\mu,\sigma^2)$ and the series converges for all t. $\left.\right]$

$$= e^{t\mu_x+s\mu_y+(1-\rho^2)\frac{t^2\sigma_x^2}{2}}\int_{-\infty}^{\infty} e^{v(\rho t\sigma_x+s\sigma_y)-v^2/2}\,dv$$

$$= e^{t\mu_x+s\mu_y+(1-\rho^2)\frac{t^2\sigma_x^2}{2}} \cdot e^{(\rho t\sigma_x+s\sigma_y)^2\cdot\frac{1}{2}}$$ [Converges for all s]

$$= e^{\mu_x t+\mu_y s+\frac{1}{2}\left[(1-\rho^2)t^2\sigma_x^2+\rho^2 t^2\sigma_x^2+2st\rho\sigma_y\sigma_x+s^2\sigma_y^2\right]}$$

$$= e^{\mu_x t+\mu_y s+\frac{1}{2}\left[t^2\sigma_x^2+s^2\sigma_y^2+2st\rho\sigma_x\sigma_y\right]}$$

Application :—

$$M_{XY}(t,s) = e^{\mu_x t + \mu_y s + \frac{1}{2}(\sigma_x^2 t^2 + \sigma_y^2 s^2 + 2\rho\sigma_x\sigma_y ts)}$$

Putting, $s=0$

$$M_{XY}(t,s) = E(e^{tX}) = M_X(t) = e^{\mu_x t + \frac{1}{2}\sigma_x^2 t^2}$$

Mgf of $N(\mu_x, \sigma_x^2)$

$\Rightarrow X \sim N(\mu_x, \sigma_x^2)$ and similarly,

$Y \sim N(\mu_y, \sigma_y^2)$

$\therefore$ Marginal distribution of X, Y are the above.

Characterisation of Bivariate Normal Distribution :—

$$Z = aX + bY$$

$$M_Z(t) = E(e^{tZ}) = E(e^{atX + btY}) = M_{XY}(at, bt)$$

$$= e^{(a\mu_x + b\mu_y)t + \frac{1}{2}t^2(a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)}$$

$$= e^{\mu_* t + \sigma_*^2 t^2/2}$$

$$\Rightarrow Z \sim N(\mu_*, \sigma_*^2)$$

$$= N(a\mu_x + b\mu_y,\ a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$$

[Converse is also true]

Recursion Relation for moments :—

Let $(X,Y) \sim BN(0,0,1,1,\rho)$

Note that, $(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$\Leftrightarrow \left(\frac{X-\mu_x}{\sigma_x}, \frac{Y-\mu_y}{\sigma_y}\right) \sim BN(0,0,1,1,\rho)$$

$$M_{XY}(t,s) = e^{\frac{1}{2}(t^2 + s^2 + 2\rho ts)}$$

$$= \sum_{r_1=0}^{\infty}\sum_{r_2=0}^{\infty} \frac{t^{r_1} s^{r_2}}{r_1!\, r_2!}\mu_{r_1, r_2}$$

$$\mu_{r_1 r_2} = E(X^{r_1} Y^{r_2}) \quad [\because E(X) = 0 = E(Y)]$$

$$\frac{\partial M}{\partial t} = (t+\rho s)M$$
$$= \sum_{r_1}\sum_{r_2} r_1 \cdot \frac{t^{r_1-1} s^{r_2}}{r_1!\, r_2!} \mu_{r_1,r_2}$$

$$\frac{\partial M}{\partial s} = (s+\rho t)\, M$$
$$= \sum_{r_1}\sum_{r_2} r_2 \cdot \frac{t^{r_1} s^{r_2-1}}{r_1!\, r_2!} \mu_{r_1 r_2}$$

$$\frac{\partial^2 M}{\partial t\, \partial s} = \rho M + (t+\rho s)(s+\rho t)\, M$$
$$= \sum_{r_1}\sum_{r_2} \frac{t^{r_1-1} s^{r_2-1}}{(r_1-1)!\,(r_2-1)!} \mu_{r_1 r_2}$$

$$\frac{\partial^2 M}{\partial t\, \partial s} - \rho t \frac{\partial M}{\partial t} - \rho s \frac{\partial M}{\partial s}$$
$$= \rho M + (t+\rho s)(s+\rho t)M - \rho t\,(t+\rho s)M - \rho s\,(s+\rho t)\,M$$
$$= \rho M - (1-\rho^2)Mts$$

$$\Rightarrow \sum_{r_1}\sum_{r_2} \frac{t^{r_1-1} s^{r_2-1}}{(r_1-1)!\,(r_2-1)!} \mu_{r_1 r_2} - \rho \sum_{r_1}\sum_{r_2} r_1 \frac{t^{r_1} s^{r_2}}{r_1!\, r_2!} \mu_{r_1 r_2}$$
$$- \rho \sum_{r_2}\sum_{r_2} r_2 \frac{t^{r_1} s^{r_2}}{r_1!\, r_2!} \mu_{r_1 r_2} = \rho \sum_{r_1}\sum_{r_2} \frac{t^{r_1} s^{r_2}}{r_1!\, r_2!} \mu_{r_1 r_2}$$
$$- (1-\rho^2) \sum_{r_1}\sum_{r_2} (r_1+1)(r_2+1) \frac{t^{r_1+1} s^{r_2+1}}{(r_1+1)!\,(r_2+1)!} \mu_{r_1 r_2}$$

Collecting coefficient of $\dfrac{t^{r_1-1} s^{r_2-1}}{(r_1-1)!\,(r_2-1)!}$

we get,

$$\mu_{r_1 r_2} - \rho(r_1-1)\mu_{r_1-1, r_2-1} - \rho(r_2-1)\mu_{r_1-1, r_2-1}$$
$$= \rho\mu_{r_1-1, r_2-1} - (1-\rho^2)(r_1-1)(r_2-1)\mu_{r_1-2, r_2-2}$$

In general, the recursion set of $BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ will be

$$\mu_{r_1 r_2} = \rho\,\sigma_x \sigma_y (r_1+r_2-1)\mu_{r_1-1, r_2-1} + (1-\rho^2)\,\sigma_x^2 \sigma_y^2 (r_1-1)\cdot (r_2-1)\mu_{r_1-2, r_2-2},$$

Example :—

$$\mu_{2,2} = \rho\sigma_x\sigma_y 3\mu_{1,1} + (1-\rho^2)\,\sigma_x^2\sigma_y^2\mu_{0,0}$$
$$= 3\rho^2\sigma_x^2\sigma_y^2 + (1-\rho^2)\,\sigma_x^2\sigma_y^2$$
$$= \sigma_x^2\sigma_y^2(1+2\rho^2)$$

Note :→ $(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$\mu_{r_1 r_2} = E(X-\mu_x)^{r_1}(Y-\mu_y)^{r_2}$$
$$= \sigma_x^{r_1}\sigma_y^{r_2}\, E\left(\frac{X-\mu_x}{\sigma_x}\right)^{r_1}\left(\frac{Y-\mu_y}{\sigma_y}\right)^{r_2}$$

$$(U,V) \sim BN(0,0,1,1,\rho) = \sigma_x^{r_1}\sigma_y^{r_2} E(U^{r_1}V^{r_2})$$

Marginal and Conditional Distribution :—

$$(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

$$f_{XY}(x,y) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right\}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma_x}\cdot e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2(1-\rho^2)}\left(u-\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right)^2}}{\sqrt{2\pi}\,\sqrt{1-\rho^2}}\,du$$
$$= \frac{1}{\sigma_x\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}, \quad x\in\mathbb{R}$$

$$u = \frac{y-\mu_y}{\sigma_y}$$
$$du = \frac{1}{\sigma_y}\,dy$$

Similarly,
$$f_Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}, \quad y\in\mathbb{R}$$

$$f_{Y/X}(y/x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y-\mu_y}{\sigma_y}\right)-\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right\}^2\right]$$

$$\therefore Y/X=x \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x),\ \sigma_y^2(1-\rho^2)\right)$$

similarly,
$$X/Y=y \sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y),\ \sigma_x^2(1-\rho^2)\right)$$

Remark:— Regression of Y on X is given by,

$$E(Y/X) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(X - \mu_x), \quad \text{linear in } X.$$

Here, least square linear regression of Y on X is identical to the true regression of Y on X.

Similar things happens for the regression of X on Y.

Note:— Though the conditional means of $Y/X=x$ differ for varying x, the distributions are homoscedastic.

[i.e. equal variences]

MGF:— $(X,Y) \sim BN(0,0,1,1,\rho)$

$$Y/X \sim N(\rho x, 1-\rho^2)$$

$$X \sim N(0,1)$$

$$M_{XY}(t,s) = E\left(e^{tX+sY}\right)$$

$$= E\left[E\left(e^{tX+sY}/X\right)\right]$$

$$= E\left(e^{tX} E\left[e^{sY}/X\right]\right)$$

$$= E\left[e^{tX} e^{\rho s + \frac{(1-\rho^2)s^2}{2}}\right]$$

$$= e^{(1-\rho^2)\frac{s^2}{2}} E\left[e^{X(t+\rho s)}\right]$$

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

$$= e^{\frac{1}{2}\left[t^2 + 2\rho t s + s^2\right]}$$

Let, $(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$\left(\frac{X-\mu_x}{\sigma_x}, \frac{Y-\mu_y}{\sigma_y}\right) \sim BN(0,0,1,1,\rho)$$

$$M_{XY}(t,s) = E\left(e^{tX+sY}\right) = e^{\mu_x t + \mu_y s} E\left[e^{\sigma_x t \frac{(x-\mu_x)}{\sigma_x} + \sigma_y s \frac{(y-\mu_y)}{\sigma_y}}\right]$$

$$= e^{\mu_x t + \mu_y s} \cdot e^{\frac{1}{2}\left(\sigma_x^2 t^2 + \sigma_y^2 s^2 + 2\rho\sigma_x\sigma_y t s\right)}$$

Result:-

$$E[g(X,Y)] = E[E\{g(X,Y)/X\}]$$

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

$$(X,Y) \sim BN(0,0,1,1,\rho)$$

$$\mu_{3,1} = E[X^3Y]$$
$$= E[E(X^3Y/X)]$$
$$= E[X^3 E(Y/X)]$$
$$= E[X^3 \rho X]$$
$$= \rho E[X^4]$$
$$= 3\rho$$

$$\mu_{2,2} = E[X^2Y^2]$$
$$= E[E(X^2Y^2/X)]$$
$$= E[X^2 E(Y^2/X)]$$
$$= E[X^2[(1-\rho^2)+\rho^2X^2]]$$
$$= (1-\rho^2)E(X^2) + \rho^2 E(X^4)$$
$$= (1-\rho^2) + 3\rho^2$$
$$= 1+2\rho^2.$$

☆ **Problem 1.** $(X,Y) \sim BN(0,0,1,1,\rho)$

Find the correlation coefficient between X and $\Phi(Y)$.

Soln.

$$X \sim N(0,1)$$
$$\Phi(Y) \sim R(0,1)$$
$$V(X) = 1, \quad V(\Phi(Y)) = \frac{1}{12}.$$

$$Cov(X,\Phi(Y)) = E(X\cdot\Phi(Y))$$
$$= E[E(X\Phi(Y)/Y)]$$
$$= E[\Phi(Y)E(X/Y)]$$
$$= E[\Phi(Y)\cdot\rho Y]$$
$$= \rho[E(Y\Phi(Y))]$$
$$= \rho\int_{-\infty}^{\infty} y\Phi(y)\,\varphi(y)\,dy$$
$$= \rho\left[-\varphi(y)\Phi(y)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\varphi^2(y)\,dy\right]$$
$$= \frac{\rho}{2\pi}\int_{-\infty}^{\infty} e^{-y^2}dy = \frac{\rho}{2\pi}\int_0^{\infty} e^{-t}t^{1/2-1}\,dt$$
$$= \frac{\rho}{2\sqrt{\pi}}$$

$$\therefore \rho_{X,\Phi(Y)} = \frac{\rho/2\sqrt{\pi}}{\sqrt{1/12}} = \rho\sqrt{\frac{3}{\pi}}$$

☆ Problem 2. $(X,Y) \sim BN(0,0,1,1,\rho)$

$p = P[XY>0]$ then show that,

$p = \cos(2\pi)$, $q = 1-p$.

Proof:- $P[XY>0] = P[X>0, Y>0] + P[X<0, Y<0]$

$= P[X>0, Y>0] + P[-X>0, -Y>0]$

$(X,Y) \sim BN(0,0,1,1,\rho)$

$(-X,-Y) \sim BN(0,0,1,1,\rho)$

$\therefore P(XY>0) = 2P[X>0, Y>0]$

$$= 2\int_0^\infty\int_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[x^2-2\rho xy+y^2\right]} dx\,dy$$

$(x,y) \longrightarrow (u,v)$

$y = uv,$

$x = u$

$x>0,\ y>0,\ \Rightarrow u>0, v>0$

$$J\left(\frac{x,y}{u,v}\right) = \begin{vmatrix} 1 & 0 \\ v & u \end{vmatrix} = u$$

$$p = P[XY>0] = \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^\infty\int_0^\infty u\cdot e^{-\frac{u^2}{2(1-\rho^2)}(v^2-2\rho v+1)} du\,dv$$

$\frac{u^2}{2} = \omega$

$u\,du = d\omega$

$$= \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^\infty \left\{\int_0^\infty e^{-\omega c}\, d\omega\right\} dv$$

$$= \frac{1}{\pi\sqrt{1-\rho^2}} \int_0^\infty \frac{dv}{\frac{v^2-2\rho v+1}{1-\rho^2}}$$

$$= \frac{\sqrt{1-\rho^2}}{\pi} \int_0^\infty \frac{dv}{(v-\rho)^2+(1-\rho^2)} = \frac{\sqrt{1-\rho^2}}{\pi\sqrt{1-\rho^2}} \tan^{-1}\left(\frac{v-\rho}{\sqrt{1-\rho^2}}\right)\Bigg|_0^\infty$$

$$= \frac{1}{\pi}\left[\frac{\pi}{2} - \tan^{-1}\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)\right]$$

$$= \frac{1}{\pi}\cot^{-1}\left(-\frac{\rho}{\sqrt{1-\rho^2}}\right) = \frac{1}{\pi}\cos^{-1}(-\rho)$$

$$\therefore -\rho = \cos p\pi$$
$$= \cos q\pi$$
$$\therefore \rho = \cos q\pi$$

☆ **Problem 3.** $(X,Y) \sim BN(0,0,1,1,\rho)$

ρ_g = correlation coefficient of $\Phi(X)$ & $\Phi(Y)$, show that,

$$\rho = 2\sin\left(\frac{\pi \rho_g}{6}\right)$$

Proof:-

$$\rho_g = \frac{E[\Phi(X)\Phi(Y)] - \frac{1}{4}}{1/12}$$

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

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P[X' \le x, Y' \le y]\, f_{XY}(x,y)\,dx\,dy, \text{ where,}$$

$(X',Y') \sim BN(0,0,1,1,\rho)$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P[X' \le X, Y' \le Y / X = x, Y = y]\; f_{XY}(x,y)\,dx\,dy$$

$$= P[X' \le X, Y' \le Y]$$

$$= P[X - X' \ge 0, Y - Y' \ge 0]$$

$$= P[U \ge 0, V \ge 0]$$, where $(U,V) \sim BN(0,0,2,2,\frac{\rho}{2})$

$$= \frac{1}{2\pi}\cos^{-1}\left(-\frac{\rho}{2}\right)$$

[Using Previous result]

$$\rho_g = \frac{\frac{1}{2\pi}\cos^{-1}\left(-\frac{\rho}{2}\right) - \frac{1}{4}}{1/12}$$

$$= \frac{6}{\pi}\cos^{-1}\left(-\frac{\rho}{2}\right) - 3$$

$$\rho = -2\cos\left[\frac{\pi}{6}(\rho_g + 3)\right]$$

$$= -2\cos\left[\frac{\pi}{2} + \pi\frac{\rho_g}{6}\right]$$

$$= 2\sin\left(\frac{\pi \rho_g}{6}\right)$$

Note,

$X, Y \longrightarrow$ iid.

$P[X > Y]$

$$= \int P[X > Y / Y = y]\, f_Y(y)\,dy$$

$$= \int P[X > y]\, f_Y(y)\,dy$$

$$= \int (1 - F(y))\, f_Y(y)\,dy$$

$$= E(1 - F(Y))$$

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

☆ Problem 4. $(X, Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ then show that,

$$E[Y/X>k] = \mu_y + \rho\sigma_y \cdot \frac{\phi\left(\frac{k-\mu_x}{\sigma_x}\right)}{1-\Phi\left(\frac{k-\mu_x}{\sigma_x}\right)}$$

Proof:-

$$E[Y/X>k] = \int_k^\infty \left\{ \int_{-\infty}^\infty y \frac{f_{XY}(x,y)}{P[X>k]} dy \right\} dx$$

$$E[Y-\mu_y/X>k] = \int_k^\infty \left\{ \int_{-\infty}^\infty (y-\mu_y) \frac{N_2(x,y;\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)}{P[X>k]} dy dx \right.$$

$$(x,y) \longrightarrow (u,v) \ni$$

$$u = \frac{x-\mu_x}{\sigma_x}, \quad v = \frac{y-\mu_y}{\sigma_y}$$

$$= \sigma_y \int_{\frac{k-\mu_x}{\sigma_x}}^\infty \left\{ \int_{-\infty}^\infty \frac{v\,BN(u,v;0,0,1,1,\rho)}{P[X>k]} dv \right\} du$$

$$= \frac{\sigma_y}{P[X>k]} \int_{\frac{k-\mu_x}{\sigma_x}}^\infty \frac{e^{-u^2/2}}{\sqrt{2\pi}} \left\{ \int_{-\infty}^\infty \frac{v e^{-\frac{1}{2}(v-\rho u)^2 \cdot \frac{1}{1-\rho^2}}}{\sqrt{2\pi}\sqrt{1-\rho^2}} dv \right\} du$$

$$= \frac{\rho\sigma_y}{P[X>k]} \int_{\frac{k-\mu_x}{\sigma_x}}^\infty u \cdot \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$$

$$= \frac{\rho\sigma_y}{P[X>k]} \left[-\phi(u)\right]_{\frac{k-\mu_x}{\sigma_x}}^\infty$$

$$= \frac{\rho\sigma_y \phi\left(\frac{k-\mu_x}{\sigma_x}\right)}{1-\Phi\left(\frac{k-\mu_x}{\sigma_x}\right)}$$

$$\therefore E[Y/X>k] = \mu_y + \rho\sigma_y \left\{ \frac{\phi\left(\frac{k-\mu_x}{\sigma_x}\right)}{1-\Phi\left(\frac{k-\mu_x}{\sigma_x}\right)} \right\}$$

Linear Transformation :— Consider the transformation of $\underset{\sim}{x} = (x_1, x_2, \dots, x_n)^T$ to $\underset{\sim}{y} = (y_1, y_2, \dots, y_n)^T$ such that

$$y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n$$
$$y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n$$
$$\vdots$$
$$y_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n$$

which is known as linear transformation.

Write $\underset{\sim}{y}_{n\times1} = A_{n\times n}\underset{\sim}{x}_{n\times1}$, where $A = [a_{ij}]_{n\times n}$

(a) If A is n.s., then $\underset{\sim}{y} = A\underset{\sim}{x}$ is called n.s. transformation.
Clearly, a n.s. transformation is one-to-one.

(b) If A is orthogonal, then $\underset{\sim}{y} = A\underset{\sim}{x}$ is called an orthogonal transformation.
Clearly, an orthogonal transformation is also an n.s. transformation

Jacobian of the linear transformation :—

$$J = \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(y_1, y_2, \dots, y_n)} = \frac{1}{\frac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)}}$$

Now,
$$\frac{\partial(y_1, y_2, y_3, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)} = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = |A|$$

Hence, $J = \frac{1}{|A|}$ or $|A^{-1}|$.

Orthogonal Transformation:— If A is orthogonal then $\underset{\sim}{y} = A\underset{\sim}{x}$ is known as an orthogonal transformation.

Here $AA^T = I_n \Rightarrow |A| = \pm 1$ and A is n.s.

Here, $\underset{\sim}{x} = A^{-1}\underset{\sim}{y} \Rightarrow$ an orthogonal transformation is one-to-one and onto.

The jacobian of the transformation is

$$J = \frac{1}{|A|} = \pm 1.$$

$$\Rightarrow |J| = 1.$$

Note that, $\underset{\sim}{y}^T\underset{\sim}{y} = (A\underset{\sim}{x})^T(A\underset{\sim}{x})$

$$= \underset{\sim}{x}^T A^T A \underset{\sim}{x} = \underset{\sim}{x}^T\underset{\sim}{x}$$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2$$

$\Rightarrow$ the length of vector is preserved by an orthogonal transformation.

An orthogonal transformation:—

$$y_1 = \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{n}}$$

$$y_2 = \frac{x_1 - x_2}{\sqrt{1.2}}$$

$$y_3 = \frac{x_1 + x_2 - 2x_3}{\sqrt{2.3}}$$

$$y_4 = \frac{x_1 + x_2 + x_3 - 3x_4}{\sqrt{3.4}}$$

$$\vdots$$

$$y_n = \frac{x_1 + x_2 + \cdots + x_{n-1} - (n-1)x_n}{\sqrt{n(n-1)}}$$

which is known as Helmert's transformation.

Sampling Distribution Arising from Univariate Normal Population :—

Let $X_1, X_2, \ldots, X_n$ be a random ~~variable~~ sample from a population following $N(\mu, \sigma^2)$.

To derive the sampling distribution of the statistic

i) the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

ii) the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Let $Z_i = \frac{X_i - \mu}{\sigma}$, $i = 1(1)n$

As $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $Z_i \overset{iid}{\sim} N(0,1)$, $i = 1(1)n$

The PDF of $(Z_1, Z_2, \ldots, Z_n)$ is

$$f(z_1, z_2, \ldots, z_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{n}\sum_{i=1}^{n} z_i^2}, \quad z_i \in R$$

Consider the following orthogonal transformation from $(Z_1, Z_2, \ldots, Z_n)$ to $(y_1, y_2, \ldots, y_n)$ such that

$$\underset{\sim}{y} = A\underset{\sim}{x}, \text{ where}$$

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \text{ is orthogonal.}$$

Note that $\underset{\sim}{y}'\underset{\sim}{y} = \underset{\sim}{z}'\underset{\sim}{z}$

$$\Leftrightarrow \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} z_i^2 \text{ and } |J| = 1.$$

The PDF of $(Y_1, Y_2, \ldots, Y_n)$ is

$$g(y_1, y_2, \ldots, y_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2}, \quad y_i \in R$$

$$= \prod_{i=1}^{n} \left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2}\right\}$$

Hence, $Y_i \overset{iid}{\sim} N(0,1)$ $\forall$ $i = (1)n$.

Now, $Y_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right) = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \sim N(0,1)$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Again, $\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} z_i^2 - Y_1^2 = \sum_{i=1}^{n} \left(\frac{x_i-\mu}{\sigma}\right)^2 - \left\{\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right\}^2$

$$= \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2\right\}$$

$$= \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}$$

$$= \frac{(n-1)s^2}{\sigma^2}$$

Hence, $\frac{(n-1)s^2}{\sigma^2} = \sum_{i=2} Y_i^2$, is the sum of squares of $(n-1)$ indebendently distributed $N(0,1)$ variables follows a χ^2-distn. with $(n-1)$ d.f.

Now, $\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} = Y_1 \sim N(0,1)$

$$\Rightarrow \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

and $\frac{(n-1)s^2}{\sigma^2} = \sum_{i=2}^{n} Y_i^2 \sim \chi^2_{n-1}$, indebendently distributed.

Hence $\bar{x}$ and s^2 are indebendently distributed.

Theorem [Cochran's theorem] :— Let $x_1, x_2, \dots, x_n$ be iid $N(0,1)$ R.V.'s. then conditional distribution of $\chi^2 = \sum_{i=1}^{n} x_i^2$ subject to $m(<n)$ L.I.N homogeneous constraints:

$$a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \dots + a_{in}x_n = 0 \quad \forall\ i=1(1)m,$$

is a χ^2-distribution with $(n-m)$ degree of freedom.

Proof:- The given linear constraints are $\underset{\sim}{a_i}'\underset{\sim}{x} = 0$, $i=1(1)m$, where, $\underset{\sim}{a_i}' = (a_{i1}, a_{i2}, \dots, a_{in})$ are L.I.N and $\underset{\sim}{x} = (x_1, x_2, \dots, x_n)$.

The above m linear constraints can be converted as $\underset{\sim}{l_i}'\underset{\sim}{x} = 0, i=1(1)m$ where $\underset{\sim}{l_i}', i=1(1)m$ are orthogonal vectors.

Now consider an orthogonal transformation,

$$Y_1 = \underset{\sim}{l_1}'\underset{\sim}{x}$$

$$\vdots$$

$$Y_m = \underset{\sim}{l_m}'\underset{\sim}{x}$$

$$Y_{m+1} = \underset{\sim}{l_{m+1}}'\underset{\sim}{x}$$

$$\vdots$$

$$Y_n = \underset{\sim}{l_n}'\underset{\sim}{x}$$

$\Rightarrow \underset{\sim}{Y}^{n\times 1} = L^{n\times n} \underset{\sim}{X}^{n\times 1}$, where

$$L = \begin{pmatrix} \underset{\sim}{l_1}' \\ \vdots \\ l_m' \\ l_{m+1} \\ \vdots \\ l_n' \end{pmatrix} \text{ is orthogonal.}$$

Clearly, $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} x_i^2$ and $|J| = 1$.

The PDF of $\underset{\sim}{X}$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}, \quad x_i \in \mathbb{R}$$

The PDF of $\underset{\sim}{Y}$ is

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2}, \quad y_i \in \mathbb{R}$$

Hence, $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(0,1)$, unconditionally.

Hence, therefore, the conditional distribution of $\chi^2 = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2$ subject to the constraints $Y_1 = \cdots = Y_m = 0$, is equivalent to unconditional distribution $\sum_{i=m+1}^{n} Y_i^2$, which is the sum of squares of $(n-m)$ independently distributed $N(0,1)$ R.V.'s, follow χ^2_{n-m}.

<u>Ex. 1.</u> If $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find the distn. of $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$.

<u>Hints:-</u>

First show that, — $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} = Y_1 \sim N(0,1)$.

Then, $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} Y_i^2 \sim \chi^2_{n-1}$

By definition of t-distn.,

$$\frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2} \Big/ n-1}} \sim t_{n-1}$$

$$\Rightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}.$$

Ex. 2. Let X_1, X_2, X_3 be iid $N(0,1)$ variates. Express the following probabilities in terms of the C.D.F. of some R.V.s.

(i) $P[2X_1 + 3X_2 + 5X_3 \geq 5]$, (ii) $P[X_1^2 + X_2^2 + X_3^2 \geq 5]$,

(iii) $P[X_1^2 - X_2^2 - X_3^2 \leq 0]$ (iv) $P[X_1^2 - 2X_2^2 - 2X_3^2 \leq 0]$,

(v) $P[4X_1^2 + 4X_2^2 - 7X_3^2 \leq 0]$. [Outline, 10.12]

Hints:-

(i) $2X_1 + 3X_2 + 5X_3 \sim N(0, 2^2 + 3^2 + 5^2)$

$\Rightarrow Y_1 \sim N(0, 38)$

$\therefore$ Required probability $= P[Y_1 \geq 5] = P\left[\frac{Y_1 - 0}{\sqrt{38}} \geq \frac{5}{\sqrt{38}}\right]$

$= 1 - \Phi\left(\frac{5}{\sqrt{38}}\right)$

(ii) $Y_2 = \sum_{i=1}^{3} X_i^2 \sim \chi^2_3$

$\therefore$ Required prob. $= P[Y_2 \geq 5] = 1 - F_{\chi^2_3}(5)$

(iii) $X_1^2 - X_2^2 - X_3^2 \leq 0$

$\Rightarrow \frac{X_1^2}{X_2^2 + X_3^2} \leq 1$

$\Rightarrow \frac{X_1^2}{\frac{X_2^2 + X_3^2}{2}} \leq 2$

$\therefore Y_3 = \frac{X_1^2/1}{\frac{X_2^2 + X_3^2}{2}} \sim F_{1,2}$

$\therefore$ Requ. Prob. $= P[Y_3 \leq 2] = P[F_{1,2} \leq 2]$

Ex.3. (a) S.T. for a r.s. $X_1, X_2, \ldots, X_n$ from $\underset{i}{N}(\mu, \sigma^2)$,

$$\sum_{i=2}^{n} \frac{i}{i-1}(X_i - \bar{X}_i)^2 \sim \sigma^2 \chi^2_{n-1}, \text{ where } \bar{X}_i = \frac{1}{i}\sum_{j=1}^{i} X_j .$$

(b) If $(X_1, X_2, \ldots, X_n)$ is a random sample from $N(\mu, \sigma^2)$, show that,

$$\frac{\sqrt{\frac{n}{n-1}}\,(X_1 - \bar{X})}{\sqrt{\dfrac{(n-1)s^2 - \frac{n}{n-1}(X_1-\bar{X})^2}{n-2}}} \sim t_{n-2}$$

Soln. → (a) Let $Y_i = \sqrt{\frac{i}{i-1}}(\bar{X}_i - X_i) = \sqrt{\frac{i}{i-1}}\left(\frac{X_1 + X_2 + \cdots + X_i}{i} - X_i\right)$

$$= \sqrt{\frac{i}{i-1}}\left(\frac{X_1 + \cdots + X_{i-1} - (i-1)X_i}{i}\right)$$

$$= \frac{X_1 + \cdots + X_{i-1} - (i-1)X_i}{\sqrt{i(i-1)}} \quad \forall\ i = 2(1)n.$$

Let, $Y_1 = \frac{1}{\sqrt{n}}X_1 + \cdots + \frac{1}{\sqrt{n}}X_n$.

Hence the transformation reduces to $\underset{\sim}{Y} = A\underset{\sim}{X}$, where

$$A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{1.2}} & -\frac{1}{\sqrt{1.2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix}$$ is orthogonal

[Helment's transformation]

The PDF of $(X_1, \ldots, X_n)$ is

$$f(x_1, \ldots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}}, \quad x_i \in \mathbb{R}$$

Here $\underset{\sim}{y}'\underset{\sim}{y} = \underset{\sim}{x}'\underset{\sim}{x}$ and $y_1 = \sqrt{n}\,\bar{x}$.

$\therefore\ |J| = 1$ is the jacobian of the transformation.

Scanned by CamScanner

The PDF of $(Y_1, \dots, Y_n)$ is

$$g(y_1, y_2, \dots, y_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n} y_i^2 - 2\mu\sqrt{n}\, y_1 + n\mu^2}{2\sigma^2}}$$

$$= \left\{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_1 - \mu\sqrt{n})^2}{2\sigma^2}}\right\} \prod_{i=2}^{n} \left\{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y_i^2}{2\sigma^2}}\right\}$$

Hence $Y_1 \sim N(\mu\sqrt{n}, \sigma^2)$ and $Y_i \sim N(0, \sigma^2)$, $i = 1(1)n$ independently distributed.

Now, $\sum_{i=2}^{n} \frac{i}{i-1} (x_i - \overline{x_i})^2 / \sigma^2$

$= \sum_{i=2}^{n} \left(\frac{Y_i}{\sigma}\right)^2$, which is the sum of squares of $(n-1)$ i.i.d. $N(0,1)$ R.V.'s, follows χ^2_{n-1}.

$$\left[Y_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 2(1)n \Rightarrow \frac{Y_i}{\sigma} \overset{iid}{\sim} N(0,1) \right]$$

$$\Rightarrow \sum_{i=2}^{n} \frac{i}{i-1} (x_i - \overline{x_i})^2 \sim \sigma^2 \chi^2_{n-1}$$

<u>Note</u>:- $Y_1 = \sqrt{n}\,\overline{X} \sim N(\mu\sqrt{n}, \sigma^2)$

$$\Rightarrow \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

(b) Consider the transformation

$\underset{\sim}{Y} = A(\underset{\sim}{X} - \mu \underset{\sim}{1})$, where

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{-(n-1)}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$ is orthogonal.

Note that $\underset{\sim}{Y}'\underset{\sim}{Y} = (\underset{\sim}{X} - \underset{\sim}{\mu})'(\underset{\sim}{X} - \underset{\sim}{\mu})$

$$\Rightarrow \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (x_i - \mu)^2$$

and $|J| = 1$.

The PDF of $\underset{\sim}{X}$ is

$$f_X(\underset{\sim}{x}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}, \quad x_i \in R$$

The PDF of $\underset{\sim}{Y}$ is

$$f_Y(\underset{\sim}{y}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2}, \quad y_i \in R$$

$$\Rightarrow Y_i \overset{iid}{\sim} N(0,\sigma^2), \quad i=1(1)n$$

Here, $Y_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) = \sqrt{n}(\bar{X}-\mu)$

and $$Y_2 = \frac{-(n-1)(X_1-\mu) + (X_2-\mu) + \cdots + (X_n-\mu)}{\sqrt{n(n-1)}}$$

$$= \frac{n\bar{X} - nX_1}{\sqrt{n(n-1)}} = -\sqrt{\frac{n}{n-1}}(X_1-\bar{X})$$

Hence, $\sqrt{\frac{n}{n-1}}(X_1-\bar{X}) = -Y_2 \sim N(0,\sigma^2)$

and $$\sum_{i=3}^{n} Y_i^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - Y_1^2 - Y_2^2$$

$$= \left\{\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right\} - \frac{n}{n-1}(X_1-\bar{X})^2$$

$$= \sum_{i=1}^{n}(X_i-\bar{X})^2 - \frac{n}{n-1}(X_1-\bar{X})^2$$

Therefore, $\sqrt{\frac{n}{n-1}}\left(\frac{X_1-\bar{X}}{\sigma}\right) = -\frac{Y_2}{\sigma} \sim N(0,1)$

and, $$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2 - \frac{n}{n-1}(X_1-\bar{X})^2}{\sigma^2} = \sum_{i=3}^{n}\left(\frac{Y_i}{\sigma}\right)^2,$$

the sum of squares of $(n-2)$ iid $N(0,1)$ R.V.'s, follows χ^2_{n-2}, independently.

By defn. of t-distn.,

$$\frac{\sqrt{\frac{n}{n-1}}\left(\frac{X_1-\bar{X}}{\sigma}\right)}{\sqrt{\left\{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2 - \frac{n}{n-1}(X_1-\bar{X})^2}{\sigma^2}\right\}/(n-2)}} \sim t_{n-2}$$

$$\Rightarrow \frac{\sqrt{\frac{n}{n-1}}(X_1-\bar{X})}{\sqrt{\left\{(n-1)s^2 - \frac{n}{n-1}(X_1-\bar{X})^2\right\}/(n-2)}} \sim t_{n-2}$$

BIVARIATE NORMAL

☆ **PROBLEM:-** Suppose $(X,Y) \sim BN(0,0,1,1,\rho)$. S.T.

i) $\dfrac{X^2 - 2\rho XY + Y^2}{1-\rho^2} \sim \chi^2_2$

ii) the m.g.f. of $Z = XY$ is

$$M_Z(t) = \left[\{1-(1+\rho)t\}\{1+(1-\rho)t\}\right]^{-1/2}$$

Solution:-

i) $(X,Y) \sim BN(0,0,1,1,\rho)$

$\therefore$ The joint PDF of (X,Y) is given by,

$$f_{XY}(x,y) = \frac{1}{\sigma_x \sigma_y \sqrt{1-\rho^2}\cdot 2\pi} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)} \quad ; (x,y) \in \mathbb{R}^2$$

Let, $U = X+Y$
$V = X-Y$

$\therefore |J| = \frac{1}{2}$, $X = \frac{U+V}{2}$, $Y = \frac{U-V}{2}$

Now, note that,

$$\frac{1}{1-\rho^2}\left\{\frac{(U+V)^2}{4} + \frac{(U-V)^2}{4} - \rho\frac{(U+V)(U-V)}{2}\right\}$$

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

$$= \frac{1}{2(1-\rho^2)}\left\{U^2 - \rho U^2 + V^2 - \rho V^2\right\}$$

$$= \frac{1}{2(1+\rho)(1-\rho)}\left\{U^2(1-\rho) + V^2(1+\rho)\right\}$$

$$= \frac{U^2}{2(1+\rho)} + \frac{V^2}{2(1-\rho)}, \longrightarrow (i)$$

$\therefore$ Joint PDF of U and V is given by,

$$f_{UV}(u,v) = \frac{1}{\sqrt{2}\sqrt{2\pi}\sqrt{1+\rho}} e^{-\frac{1}{4(1+\rho)}u^2} \cdot \frac{1}{\sqrt{2}\sqrt{2\pi}\sqrt{1-\rho}} e^{-\frac{1}{4(1-\rho)}v^2}, \quad (u,v) \in \mathbb{R}^2$$

$\therefore$ U and V are independent.

$U \sim N(0, 2(1+\rho))$

$\therefore \dfrac{U}{\sqrt{2(1+\rho)}} \sim N(0,1) \Rightarrow \dfrac{U^2}{2(1+\rho)} \sim \chi^2_1$

Similarly, $\dfrac{V^2}{2(1-\rho)} \sim \chi^2_1$.

$\therefore \dfrac{U^2}{2(1+\rho)} + \dfrac{V^2}{2(1-\rho)} \sim \chi^2_2$ [By the reproductive property of χ^2- distribution]

ii) MGF of X, Y is given by,

$$M_{XY}(t) = E\left(e^{tXY}\right)$$

$$= E\left[E\left(e^{tXY} \middle| X\right)\right]$$

$$Y|X \sim N\left(\rho x, (1-\rho^2)\right)$$

$$\therefore E\left[e^{tX \cdot \rho x + \frac{1}{2}t^2 x^2 (1-\rho^2)}\right]$$

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

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

$$= \frac{1}{\left[1-2\left(t\rho + \frac{1}{2}t^2(1-\rho^2)\right)\right]^{1/2}} \quad [\because X^2 \sim \chi^2_1]$$

$$= \frac{1}{\left[1-2t\rho - t^2(1-\rho^2)\right]^{1/2}}$$

$$= \frac{1}{\sqrt{(1-t\rho)^2 - t^2}}$$

$$= \frac{1}{\sqrt{\{(1-\rho)t+1\}\{1-t(1+\rho)\}}}$$

$$= \left[\{1-(1+\rho)t\}\{1+(1-\rho)t\}\right]^{-1/2}$$

Bivariate Normal Distribution

Definition:— A bivariate R.V. (X, Y) is said to have a bivariate normal distribution if the PDF of (X, Y) is of the following form:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}, \quad (x,y) \in \mathbb{R}^2$$

where $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $|\rho| < 1$.

Then, we write $(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

Marginal Distribution:—

Note that, $\frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}$

$$= \left[\frac{\left\{y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}^2}{\sigma_2^2(1-\rho^2)} + \left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]$$

$$= \frac{(y-\beta x)^2}{\sigma_{2.1}^2} + \frac{(x-\mu_1)^2}{\sigma_1^2}, \text{ where } \beta x = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \quad \sigma_{2.1}^2 = \sigma_2^2(1-\rho^2)$$

Marginal PDF of X is,

$$f_X(x) = \frac{\exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]}{\sigma_1\sqrt{2\pi}} \times \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\cdot\frac{(y-\beta x)^2}{\sigma_{2.1}^2}}}{(\sqrt{2\pi})\,\sigma_{2.1}}\,dy$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}, \quad x \in \mathbb{R}$$

Here, $X \sim N(\mu_1, \sigma_1^2)$

similarly, it can be shown that $Y \sim N(\mu_2, \sigma_2^2)$.

Conditional Distribution:— The PDF of Y given $X = x$ is

$$f_{Y/X}(y/x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\beta x}{\sigma_{2.1}}\right)^2\right\}}}{\frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}$$

$$= \frac{1}{\sigma_{2.1}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta x}{\sigma_{2.1}}\right)^2}, \quad y \in \mathbb{R}$$

Hence, $Y/X=x \sim N(\beta x, \sigma_{2.1}^2)$

$\Leftrightarrow Y/X=x \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ \sigma_2^2(1-\rho^2)\right)$

similarly, it can be shown that,

$$X/Y=y \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2),\ \sigma_1^2(1-\rho^2)\right)$$

Remark :-

(1) Note that $E(Y/X=x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x-\mu_1)$ and
$Var(Y/X=x) = \sigma_2^2(1-\rho^2)$

Hence, the regression of Y on X is linear and conditional distn. is homo-scedastic.

(2) $E(XY) = E[E(XY/X)] = E[X.E(Y/X)]$

$$= E\left[X\left\{\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}\right]$$

$$= \mu_1\mu_2 + \rho\frac{\sigma_2}{\sigma_1}.\sigma_1^2 \quad \left[\because E\{X(X-\mu_1)\} = E(X-\mu_1)^2 = \sigma_1^2\right]$$

$$\Rightarrow E(XY) = \mu_1\mu_2 + \rho\sigma_1\sigma_2$$

$$\Rightarrow \frac{E(XY)-\mu_1\mu_2}{\sigma_1\sigma_2} = \rho.$$

$$\Rightarrow \rho_{XY} = \rho.$$

(3) If $\rho^2 = 1$, then the PDF becomes undefined.
But $\rho = \pm 1$, then $P[\alpha X + \beta Y + \gamma = 0] = 1$.
for some non-null (α, β), which is known as Singular or degenerate Bivariate distribution.

Ex. 1. Consider the bivariate PDF :

$$f(x,y) = \frac{1}{2}\left[\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{x^2-2\rho xy+y^2\}} + \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\left\{\frac{x^2+2\rho xy+y^2}{2(1-\rho^2)}\right\}}\right]$$

$$= \frac{1}{2}\left[\phi(x,y;0,0,1,1,\rho) + \phi(x,y;0,0,1,1,-\rho)\right]; (x,y)\in\mathbb{R}^2$$

Show that the marginal distn. are normals and comment on the fact. Also, S.T. $\rho_{XY} = 0$ and comment.

Solution:- Note that $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$$= \frac{1}{2}\left[\frac{1}{\sqrt{2\pi}}e^{-x^2/2} + \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right]$$

$$= \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad x\in\mathbb{R}$$

and $f_Y(y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}, \quad y\in\mathbb{R}$

Here, the marginal distns are univariate normals but the joint PDF $f(x,y)$ is not a bivariate normal PDF.

Now, $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - 0.0}{1.1} = E(XY)$

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

$$= \frac{1}{2}\Big[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, \phi(x,y;0,0,1,1,\rho)\, dx\, dy$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, \phi(x,y;0,0,1,1,-\rho)\, dx\, dy \Big]$$

$$= \frac{1}{2}[\rho + (-\rho)] = 0.$$

[Here $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, \phi(x,y;0,0,1,1,\rho)\, dx\, dy$ is the correlation coefficient of the R.V.'s whose joint distn is $BN(0,0,1,1,\rho)$ and it is ρ]

Hence, $\rho_{XY} = 0$ but $f(x,y) \neq f_X(x) \cdot f_Y(y)$

$\Rightarrow \rho_{XY} = 0$ but the RV's X and Y are not independent.

MGF of Bivariate Normal Distribution : —

Let $(X,Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

Define, $U = \frac{X-\mu_1}{\sigma_1}$, $V = \frac{Y-\mu_2}{\sigma_2}$

Then $(U,V) \sim BN(0,0,1,1,\rho)$

Now, $M_{U,V}(t_1,t_2) = E[e^{t_1U + t_2V}]$

$$= E\{E[e^{t_1U + t_2V}/U]\} = E\{e^{t_1U} E[e^{t_2V}/U]\}$$

$$= E\{e^{t_1U} . M_{V/U}(t_2)\}$$

[N.T. $V/U=u \sim N(\rho u, (1-\rho^2))$ & $U \sim N(0,1)$]

$$= E\left[e^{t_1U} . e^{t_2\rho u + \frac{1}{2}t_2^2(1-\rho^2)}\right]$$

$$= e^{\frac{1}{2}t_2^2(1-\rho^2)} E\left[e^{(t_1+\rho t_2)U}\right]$$

$$= e^{\frac{1}{2}t_2^2(1-\rho^2)} . e^{\frac{1}{2}(t_1+\rho t_2)^2}$$

$$= e^{\frac{1}{2}\{t_1^2 + t_2^2 + 2\rho t_1 t_2\}}$$

Now, $M_{X,Y}(t_1,t_2)$

$$= E[e^{t_1X + t_2Y}] = E\left[e^{t_1(\mu_1 + \sigma_1 U) + t_2(\mu_2 + \sigma_2 V)}\right]$$

$$= e^{t_1\mu_1 + t_2\mu_2} E\left[e^{(t_1\sigma_1)U + (t_2\sigma_2)V}\right]$$

$$= e^{t_1\mu_1 + t_2\mu_2} M_{U,V}(t_1\sigma_1, t_2\sigma_2)$$

$$= e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}\{t_1^2\sigma_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + t_2^2\sigma_2^2\}}$$

Sampling Distribution Arising from Bivariate Normal Distribution:

Let (X_i, Y_i), $i=1(1)n$, be a random sample from $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

Define, $U_i = Y_i - \mu_2 - \beta(X_i - \mu_1) \sim N(0, \sigma_{2\cdot1}^2)$

and, $V_i = (X_i - \mu) \sim N(0, \sigma_1^2)$ $\forall i$, independently, where

$$\beta = \rho\frac{\sigma_2}{\sigma_1}, \quad \sigma_{2\cdot1}^2 = \sigma_2^2(1-\rho^2)$$

Joint distn.s of means:—

As $U_i \overset{iid}{\sim} N(0, \sigma_{2\cdot1}^2)$, $i=1(1)n$,

$\bar{U} \sim N\left(0, \frac{\sigma_{2\cdot1}^2}{n}\right)$

As $V_i \overset{iid}{\sim} N(0, \sigma_1^2)$, $i(1)n$

$\bar{V} \sim N\left(0, \frac{\sigma_1^2}{n}\right)$

independently.

The PDF of $(\bar{U}, \bar{V})$ is $= c.\exp\left[-\frac{1}{2}\left\{\frac{n\bar{u}^2}{\sigma_{2\cdot1}^2} + \frac{n\bar{v}^2}{\sigma_1^2}\right\}\right]$, $(\bar{u}, \bar{v}) \in \mathbb{R}^2$.

Now, $\bar{u} = \bar{y} - \mu_2 - \beta(\bar{x} - \mu_1)$

$\bar{v} = \bar{x} - \mu_1$,

and $J = \frac{\partial(\bar{u}, \bar{v})}{\partial(\bar{x}, \bar{y})} = \begin{vmatrix} -\beta & 1 \\ 1 & 0 \end{vmatrix} = -1$.

The PDF of $(\bar{x}, \bar{y})$ is

$$c.\exp\left[-\frac{n}{2}\left\{\frac{(\bar{y} - \mu_2 - \beta(\bar{x} - \mu_1))^2}{\sigma_{2\cdot1}^2} + \frac{(\bar{x} - \mu_1)^2}{\sigma_2^2}\right\}\right]$$

$$= c.\exp\left[-\frac{n}{2(1-\rho^2)}\left\{\left(\frac{\bar{x} - \mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\bar{x} - \mu_1}{\sigma_1}\right)\left(\frac{\bar{y} - \mu_2}{\sigma_2}\right) + \left(\frac{\bar{y} - \mu_2}{\sigma_2}\right)^2\right\}\right]$$

Joint Distribution of Variances & co-variances:—

For given $v_1, v_2, \ldots, v_n$, $U_i \overset{iid}{\sim} N(0, \sigma_{2\cdot1}^2)$, $\forall i=1(1)n$ and define

$$S_{uu} = \sum_{i=1}^{n}(U_i - \bar{U})^2, \quad S_{uv} = \sum_{i=1}^{n} U_i(v_i - \bar{v}) \text{ and } S_{vv} = \sum_{i=1}^{n}(v_i - \bar{v})^2$$

Consider the orthogonal transformation,

$\underset{\sim}{W} = L\underset{\sim}{U}$, where $L = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{v_1 - \bar{v}}{\sqrt{S_{vv}}} & \frac{v_2 - \bar{v}}{\sqrt{S_{vv}}} & \cdots & \frac{v_n - \bar{v}}{\sqrt{S_{vv}}} \\ l_{31} & l_{32} & \cdots & l_{3n} \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{bmatrix}$ is orthogonal.

Here $|J| = 1$ and $\sum_{i=1}^{n} W_i^2 = \sum_{i=1}^{n} U_i^2$

For given $v_1, \dots, v_n$;

(i) The PDF of $\underset{\sim}{U}$ is $\quad c \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{u_i^2}{\sigma_{2\cdot1}^2}\right]$

(ii) The PDF of $\underset{\sim}{W}$ is $\quad c \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{w_i^2}{\sigma_{2\cdot1}^2}\right]$

and hence $w_i \overset{iid}{\sim} N(0, \sigma_{2\cdot1}^2)$ $\forall i = 1(1)n$.

Now, $w_1 = \sqrt{n}\,\bar{U}$

$$w_2 = \frac{\sum_{i=1}^{n} U_i(v_i - \bar{v})}{\sqrt{S_{vv}}} = \frac{S_{uv}}{\sqrt{S_{vv}}}$$

and $$\sum_{i=3}^{n} w_i^2 = \sum_{i=1}^{n} U_i^2 - n\bar{U}^2 - \frac{S_{uv}^2}{S_{vv}}$$
$$= S_{uu} - \frac{S_{uv}^2}{S_{vv}}$$

Hence, for given $v_1, \dots, v_n$;

$$\frac{S_{uv}}{\sigma_{2\cdot1}\sqrt{S_{vv}}} = \frac{w_2}{\sigma_{2\cdot1}} \sim N(0,1)$$

$$\frac{1}{\sigma_{2\cdot1}^2}\left(S_{uv} - \frac{S_{uv}^2}{S_{vv}}\right) = \sum_{i=3}^{n}\frac{w_i^2}{\sigma_{2\cdot1}^2} \sim \chi^2_{n-2}, \text{ independently.}$$

If $(v_1, v_2, \dots, v_n)$ is now considered as a R.V., then $(v_1, v_2, \dots, v_n)$ is independent of $(U_1, U_2, \dots, U_n)$ and hence $(v_1, v_2, \dots, v_n)$ is independently distributed with $(v_1, v_2, \dots, v_n)$ is

$$T_1 = \frac{S_{uv}}{\sigma_{2\cdot1}\sqrt{S_{vv}}} \sim N(0,1) \text{ and}$$

$$T_2 = \frac{1}{\sigma_{2\cdot1}^2}\left(S_{uu} - \frac{S_{uv}^2}{S_{vv}}\right) \sim \chi^2_{n-2}$$

As $v_i \overset{i.i.d.}{\sim} N(0, \sigma_1^2)$, $T_3 = \frac{S_{vv}}{\sigma_1^2} \sim \chi^2_{n-1}$.

The PDF of (T_1, T_2, T_3) is

$$= c \cdot e^{-\frac{1}{2}\{T_1^2 + T_2 + T_3\}} \cdot T_2^{\frac{n-4}{2}} \cdot T_3^{\frac{n-3}{2}}$$

Now, $$T_1 = \frac{S_{uv}}{\sigma_{2\cdot1}\sqrt{S_{vv}}} = \frac{S_{xy} - \beta S_{xx}}{\sigma_{2\cdot1}\sqrt{S_{xx}}}$$

$$\left[\because \; S_{vv} = \sum (x_i - \bar{x})^2 = S_{xx} \right.$$
$$S_{uv} = \sum_{i=1}^{n}\{y_i - \mu_2 - \beta(x_i - \mu_1)\}(x_i - \bar{x})$$
$$\left. = S_{xy} - \beta S_{xx} \right]$$

$$T_2 = \frac{S_{uu} - \frac{S_{uv}^2}{S_{vv}}}{\sigma_{2\cdot1}^2} = \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right)\cdot\frac{1}{\sigma_{2\cdot1}^2}$$

and $$T_3 = \frac{S_{vv}}{\sigma_1^2} = \frac{S_{xx}}{\sigma_1^2}$$

Note that, $$\frac{\partial(T_1, T_2, T_3)}{\partial(S_{xx}, S_{xy}, S_{yy})} = \frac{1}{\sigma_1^2}\cdot\frac{1}{\sigma_{2\cdot1}\sqrt{S_{xx}}}\cdot\frac{1}{\sigma_{2\cdot1}^2}$$

and $T_1^2 + T_2 + T_3$

$$= \frac{1}{1-\rho^2}\left\{\frac{S_{xx}}{\sigma_{2\cdot1}^2} - 2\rho\frac{S_{xx}}{\sigma_1\sigma_2} + \frac{S_{yy}}{\sigma_2^2}\right\}$$

The PDF of (S_{xx}, S_{xy}, S_{yy}) is

$$c\cdot\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{S_{xx}}{\sigma_1^2} - \frac{2\rho S_{xx}}{\sigma_1\sigma_2} + \frac{S_{yy}}{\sigma_2^2}\right\}\right]\cdot\left(S_{xx}\cdot S_{yy} - S_{xy}^2\right)^{\frac{n-4}{2}}$$

Sampling Distribution of sample correlation coefficient (r) when $\rho = 0$:—

Let $(X_i, Y_i)\ \forall\ i = 1(1)n$ be a random sample from a $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho = 0)$ distribution.

Hence, $X_i \sim N(\mu_1, \sigma_1^2),\ i = 1(1)n$
$Y_i \sim N(\mu_2, \sigma_2^2),\ i = 1(1)n$ independently.

For fixed $X_1, X_2, \ldots, X_n$; let $\underset{\sim}{V} = C(\underset{\sim}{Y} - \mu_2\underset{\sim}{1})$,

where
$$C = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{X_1-\bar{X}}{\sqrt{S_{xx}}} & \frac{X_2-\bar{X}}{\sqrt{S_{xx}}} & \cdots & \frac{X_n-\bar{X}}{\sqrt{S_{xx}}} \\ c_{31} & c_{32} & \cdots & c_{3n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$ is orthogonal.

Here $|J| = 1$ and $\sum_{i=1}^{n} V_i^2 = \sum_{i=1}^{n}(Y_i - \mu_2)^2$

For fixed $x_1, x_2, \ldots, x_n$,

i) the PDF of $\underset{\sim}{Y}$ is $c\cdot\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i-\mu_2)^2}{\sigma_2^2}\right]$

ii) the PDF of $\underset{\sim}{V}$ is $c\cdot\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{V_i^2}{\sigma_2^2}\right]$

and hence $V_i \overset{i.i.d.}{\sim} N(0, \sigma_2^2),\ i = 1(1)n.$

Now, $Y_1 = \sqrt{n}\,(\bar{Y} - \mu_2)$

$$Y_2 = \frac{\sum (x_i - \bar{x})(Y_i - \mu_2)}{\sqrt{S_{xx}}} = \frac{S_{xy}}{\sqrt{S_{xx}}} = r\sqrt{S_{yy}}$$

and $$\sum_{i=3}^{n} Y_i^2 = \left\{\sum_{i=1}^{n}(Y_i-\mu_2)^2 - n(\bar{Y}-\mu_2)^2\right\} - r^2 S_{yy}$$

$$= \sum (Y_i - \bar{Y})^2 - r^2 S_{yy}$$

$$= (1-r^2) S_{yy}$$

For fixed $X_1, X_2, \ldots, X_n$,

$$\frac{r\sqrt{S_{yy}}}{\sigma_2} = \frac{Y_2}{\sigma_2} \sim N(0,1) \text{ and}$$

$$\frac{(1-r^2)S_{yy}}{\sigma_2^2} = \sum_{i=3}^{n} \frac{Y_i^2}{\sigma_2^2} \sim \chi^2_{n-2}, \text{ independently.}$$

Now, consider $X_1, \ldots, X_n$ as a R.V., then $(X_1, X_2, \ldots, X_n)$ is independent of $(Y_1, \ldots, Y_n)$; hence $(X_1, \ldots, X_n)$ is independently distributed with $Y_2 = r\sqrt{S_{yy}}$ and $\sum_{i=3}^{n} Y_i = (1-r^2) S_{yy}$.

Hence, for a r.s. $\{(X_i, Y_i) : i = 1(1)n\}$ from the population,

$$\left.\begin{array}{l} \dfrac{r\sqrt{S_{yy}}}{\sigma_2} \sim N(0,1) \\[2ex] \dfrac{(1-r^2)S_{yy}}{\sigma_2^2} \sim \chi^2_{n-2} \end{array}\right\} \text{independently.}$$

$$\Rightarrow \frac{\dfrac{r\sqrt{S_{yy}}}{\sigma_2}}{\sqrt{\dfrac{(1-r^2)S_{yy}}{\sigma_2^2}\Big/ n-2}} \sim t_{n-2}, \text{ by definition of } t\text{-distribution.}$$

$$\Rightarrow t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}.$$

The PDF of t is $$\frac{1}{\sqrt{n-2}\cdot\beta\left(\frac{1}{2}, \frac{n-2}{2}\right)}\left(1+\frac{t^2}{n-2}\right)^{-\left(\frac{n-1}{2}\right)}, \quad t \in \mathbb{R}$$

Here, $$1 + \frac{t^2}{n-2} = \frac{1}{1-r^2}$$

$$\Rightarrow \frac{2t\,dt}{n-2} = \frac{2r\,dr}{(1-r^2)^2} \Rightarrow dt = \frac{\sqrt{n-2}\,dr}{(1-r^2)^{3/2}}$$

the PDF of r is, $$\begin{cases} \dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{\frac{n-4}{2}}, & -1 < r < 1 \\[2ex] 0 & , \text{ow} \end{cases}$$

Sampling Distribution related to Sample Linear Regression :-

(I) Non-stochastic independent variable :-

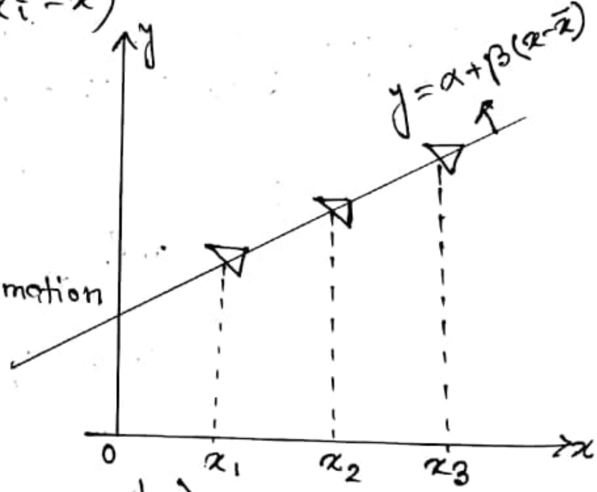
Here, we shall considered the (least squares) linear regression of Y on X. Let $x_1, x_2, \ldots, x_n$; $n \geq 3$ be the values of X.

Further assume, for $X = x_i$, $Y \sim N(\eta_{x_i}, \sigma^2)$ $\forall$ $i = 1(1)n$, where, $\eta_{x_i} = \alpha + \beta(x_i - \bar{x})$, when $x = x_i$, let Y_i be a r.s. from Y. Then $Y_i \sim N(\alpha + \beta(x_i - \bar{x}), \sigma^2)$ $\forall$ $i = 1(1)n$, independently.

By method of least squares, the samples least square linear regression of Y on X is

$$Y_x = a + b(x_i - \bar{x}) \text{, where,}$$

$$a = \bar{Y}, \quad b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$



Let $Z_i = (Y_i - \eta_{x_i})$, $i = 1(1)n$

Clearly, $Z_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1(1)n$

Consider the orthogonal transformation : $\underset{\sim}{U} = L\underset{\sim}{Z}$, where

$$L = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{x_1 - \bar{x}}{\sqrt{S_{xx}}} & \frac{x_2 - \bar{x}}{\sqrt{S_{xx}}} & \cdots & \frac{x_n - \bar{x}}{\sqrt{S_{xx}}} \\ l_{31} & l_{32} & \cdots & l_{3n} \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \text{ is orthogonal.}$$

Note that $\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n} Z_i^2$ and $|J| = 1$.

The PDF of $\underset{\sim}{Z}$ is $\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} Z_i^2}$

The PDF of $\underset{\sim}{U}$ is $\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} u_i^2}$

Clearly, $U_i \overset{iid}{\sim} N(0, \sigma^2)$ $\forall$ $i = 1(1)n$.

Note that, $U_1 = \sqrt{n}\,\bar{Z} = \sqrt{n}\,(\bar{Y}-\alpha) \sim N(0,\sigma^2)$

$$U_2 = \frac{\sum (x_i-\bar{x}) z_i}{\sqrt{S_{xx}}}$$

$$= \frac{\sum (x_i-\bar{x}) Y_i - \beta \sum (x_i-\bar{x})^2}{\sqrt{S_{xx}}}$$

$$= \frac{S_{xy} - \beta S_{xx}}{\sqrt{S_{xx}}}$$

$$= (b-\beta)\sqrt{S_{xx}} \quad \text{as } b = \frac{S_{xy}}{S_{xx}}.$$

and $\sum_{i=3}^{n} U_i^2 = \sum_{i=1}^{n} z_i^2 - n(\bar{Y}-\alpha)^2 - (b-\beta)^2 S_{xx}$

$$= \sum \{Y_i - \alpha - \beta(x_i-\bar{x})\}^2 - n(\bar{Y}-\alpha) - (b-\beta)^2 S_{xx}$$

$$= \sum_{i=1}^{n} \{Y_i - a - b(x_i-\bar{x})\}^2 \sim \sigma^2 \chi^2_{n-2}, \text{ independently}$$

$$\Big[\ \sum_{i=1}^{n} \{Y_i - \alpha - \beta(x_i-\bar{x})\}^2$$

$$= \sum_{i=1}^{n} \{Y_i - a - b(x_i-\bar{x}) + (a-\alpha) + (b-\beta)(x_i-\bar{x})\}^2$$

$$= \sum_{i=1}^{n} \{Y_i - a - b(x_i-\bar{x})\}^2 + n(\bar{Y}-\alpha)^2 + (b-\beta)^2 S_{xx}\}$$

due to normal equation. $\Big]$

Hence, $a = \bar{Y} \sim N\left(\alpha, \frac{\sigma^2}{n}\right)$, $b \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right)$ and

$$\sum_{i=1}^{n} \frac{\{Y_i - a - b(x_i-\bar{x})\}^2}{\sigma^2} \sim \chi^2_{n-2}, \text{ independently.}$$

<u>Remark:</u>– (1) $\sum_{i=1}^{n} \{Y_i - a - b(x_i-\bar{x})\}^2 = \sum_{i=1}^{n} e_i^2$ is the RSS/SSE due to the use LS linear regression as a predicting formula.

Here $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \{Y_i - a - b(x_i-\bar{x})\}^2 = S^2_{y.x}$ is an UE σ^2.

(2)
$$\frac{\frac{\sqrt{n}(a-\alpha)}{\sigma}}{\sqrt{\frac{(n-2)S^2_{y.x}}{\sigma^2}\Big/(n-2)}} \sim t_{n-2} \Rightarrow \frac{\sqrt{n}(a-\alpha)}{S_{y.x}} \sim t_{n-2}.$$

and,
$$\frac{\frac{(b-\beta)\sqrt{S_{xx}}}{\sigma}}{\sqrt{\frac{(n-2)S^2_{y.x}}{\sigma^2}\Big/n-2}} \sim t_{n-2} \Rightarrow \frac{(b-\beta)\sqrt{S_{xx}}}{S_{y.x}} \sim t_{n-2}.$$

(II). Stochastic independent variable :—

Assume that both the variables X and Y are stochastic and $(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$,

Then $Y/X = x \sim N(\eta_x, \sigma_{2.1}^2)$, where $\eta_x = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$

$= \mu_2 + \beta(x - \mu_1)$, where $\beta = \rho\frac{\sigma_2}{\sigma_1}$

and $\sigma_{2.1}^2 = \sigma_2^2(1-\rho^2)$

Let (X_i, Y_i), $i = 1(1)n$, $n \geq 3$ be a r.s. from the population.

The LS linear regression of Y on X is

$$Y_x = a + b(x - \bar{x}), \text{ where, } a = \bar{Y},\ b = \frac{S_{xy}}{S_{xx}}$$

For fixed $X_1, X_2, \ldots, X_n$

$$Y_i / X_i \sim N(\eta_{x_i}, \sigma_{2.1}^2)$$

and consider the orthogonal transformation:

$$\underset{\sim}{U} = L\begin{pmatrix} Y_1 - \eta_{X_1} \\ Y_2 - \eta_{X_2} \\ \vdots \\ Y_n - \eta_{X_n}\end{pmatrix}, \text{ where, } L = \begin{pmatrix} \frac{X_1 - \bar{X}}{\sqrt{S_{xx}}} & \frac{X_2 - \bar{X}}{\sqrt{S_{xx}}} & \cdots & \frac{X_n - \bar{X}}{\sqrt{S_{xx}}} \\ l_{21} & l_{22} & \cdots & l_{2n} \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn}\end{pmatrix}$$

is orthogonal $(\perp)$.

Here, $\sum U_i^2 = \sum (Y_i - \eta_{x_i})^2$ and $|J| = 1$

For fixed/given $x_1, x_2, \ldots, x_n$,

(i) the PDF of $\underset{\sim}{Y}$ is $c.\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - \eta_{x_i})^2}{\sigma_{2.1}^2}\right]$

(ii) the PDF of $\underset{\sim}{U}$ is $c.\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{U_i^2}{\sigma_{2.1}^2}\right]$

and hence $U_i \overset{iid}{\sim} N(0, \sigma_{2.1}^2)$, $i = 1(1)n$

Now, $$U_1 = \frac{\sum(x_i - \bar{x})(Y_i - \eta_{x_i})}{\sqrt{S_{xx}}} = (b - \beta)\sqrt{S_{xx}} \sim N(0, \sigma_{2.1}^2)$$

for fixed $x_1, x_2, \ldots, x_n$, i.e. for fixed S_{xx}.

Hence, the conditional PDF of 'b' given S_{xx} is

$$N\left(\beta, \frac{\sigma_{2.1}^2}{S_{xx}}\right).$$

Again, $$\frac{S_{xx}}{\sigma_1^2} \sim \chi^2_{n-1}.$$

Joint PDF of b and S_{xx} is

$$c.\sqrt{S_{xx}}\,.\,e^{-\frac{1}{2}\left\{\frac{S_{xx}(b-\beta)^2}{\sigma_{2.1}^2}+\frac{S_{xx}}{\sigma_1^2}\right\}}\,.\,(S_{xx})^{\frac{n-3}{2}}$$

The PDF of b is

$$c\int_0^\infty e^{-\frac{S_{xx}}{2\sigma_1^2}\left\{\frac{(b-\beta)^2\sigma_1^2}{\sigma_{2.1}^2}+1\right\}}\,(S_{xx})^{\frac{n-2}{2}}\,d(S_{xx})$$

$$= c.\frac{1}{\left\{\frac{(b-\beta)^2\sigma_1^2}{\sigma_{2.1}^2}+1\right\}^{n/2}}$$

Hence $\quad \frac{(b-\beta)\sigma_1}{\sigma_{2.1}}\sqrt{n-1} \sim t_{n-1}$.

Result:- Let $X_1, X_2, \ldots, X_n$ be independently distributed $N(\mu_i, \sigma_i^2)$, $i=1(1)n$, respectively.

Define, $U = \sum_{i=1}^{n} a_i X_i = \underset{\sim}{a}'\underset{\sim}{X}$

$V = \sum_{i=1}^{n} b_i X_i = \underset{\sim}{b}'\underset{\sim}{X}$

Then $(U,V) \sim$ BN distn.

Proof:-

$$M_{U,V}(t_1,t_2) = E\left[e^{t_1U+t_2V}\right]$$

$$= E\left[e^{\sum_{i=1}^{n}(t_1a_i+t_2b_i)X_i}\right]$$

$$= \prod_{i=1}^{n} E\left[e^{s_iX_i}\right], \text{ where } s_i = t_1a_i+t_2b_i$$

$$= \prod_{i=1}^{n} M_{X_i}(s_i)$$

$$= \prod_{i=1}^{n} e^{s_i\mu_i+\frac{1}{2}s_i^2\sigma_i^2}$$

$$= e^{\sum s_i\mu_i+\frac{1}{2}\sum s_i^2\sigma_i^2}$$

$$= e^{t_1\left(\sum_{i=1}^{n}a_i\mu_i\right)+t_2\left(\sum b_i\mu_i\right)+\frac{1}{2}\left\{t_1^2\left(\sum a_i^2\sigma_i^2\right)+t_2^2\left(\sum b_i^2\sigma_i^2\right)+2t_1t_2\sum a_ib_i\sigma_i^2\right\}}$$

which is the MGF of $BN\left(\sum a_i\mu_i, \sum b_i\mu_i, \sum a_i^2\sigma_i^2, \sum b_i^2\sigma_i^2, \rho\right)$,

where, $\rho = \dfrac{\sum_{i=1}^{n} a_ib_i\sigma_i^2}{\sqrt{\sum a_i^2\sigma_i^2}\sqrt{\sum b_i^2\sigma_i^2}}$.

Hence the proof.

Remark:- If X_i's are independently normals, then their two linear combinations jointly follows a Bivariate normal distribution.

Ex.1. "If $X_i \sim N(\mu_i, \sigma_i^2)$, $i=1(1)n$, independently then $\underset{\sim}{a}'\underset{\sim}{X}$ and $\underset{\sim}{b}'\underset{\sim}{X}$ are independently distributed iff $\underset{\sim}{a}'\underset{\sim}{b} = 0$" - Prove it or rectify it.

Hints:- $(\underset{\sim}{a}'\underset{\sim}{X}, \underset{\sim}{b}'\underset{\sim}{X})$ follows Bivariate Normal with co-variance $(\sum a_i b_i \sigma_i^2)$.

$\underset{\sim}{a}'\underset{\sim}{X}$ and $\underset{\sim}{b}'\underset{\sim}{X}$ are independent iff $\sum a_i b_i \sigma_i^2 = 0$.

If $\sigma_i^2 = \sigma^2 \ \forall i$, then $\underset{\sim}{a}'\underset{\sim}{X}$ and $\underset{\sim}{b}'\underset{\sim}{X}$ are independent iff $\sum a_i b_i = 0$ iff $\underset{\sim}{a}'\underset{\sim}{b} = \underset{\sim}{0}$.

———— X ————

Scanned by CamScanner

ORDER STATISTICS

Introduction: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n drawn from a population with distribution function F. If the observations $X_1, X_2, \ldots, X_n$ are arranged in increasing order of magnitude then the re-arranged random variables $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ are called the order statistics of the sample.

In case of sampling from continuous population we have $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ with probability 1. It is clear that the order statistic $X_{(i)}$'s are dependent though the original observations $X_1, X_2, \ldots, X_n$ are independent.

Thus the rth order statistic of the sample of size n is simply the rth order statistic (rth smallest observation in the sample) and is denoted by $X_{(r)}$.

Exact sampling distribution of order statistic :

i) Distribution of $X_{(1)}$: Let us consider a random sample $X_1, X_2, \ldots, X_n$ drawn from a population having distribution function $F(\cdot)$.

$X_{(1)}$ is the first order statistic, the distn. function of $X_{(1)}$ is given by

$$F_{X_{(1)}}(x) = 1 - P[X_{(1)} > x]$$
$$= 1 - P[X_1 > x, X_2 > x, \ldots, X_n > x]$$
$$= 1 - \prod_{i=1}^{n} P[X_i > x] \quad [\because X_1, \ldots, X_n \text{ are independent}]$$
$$= 1 - \{P[X_1 > x]\}^n$$
$$= 1 - (1 - F(x))^n$$

$\therefore$ The pdf of $X_{(1)}$ is given by,

$$f_{X_{(1)}}(x) = n(1 - F(x))^{n-1} f(x)$$

ii) Distribution of $X_{(n)}$: The distn. function of $X_{(n)}$ is given by,

$$F_{X_{(n)}}(x) = P[X_{(n)} \le x]$$
$$= P[X_1 \le x, X_2 \le x, \ldots, X_n \le x]$$
$$= (P[X_1 \le x])^n \quad [\because X_i\text{'s are i.i.d.}]$$
$$= [F(x)]^n$$

$\therefore$ The pdf of $X_{(n)}$ is given by,

$$f_{X_{(n)}}(x) = n\{F(x)\}^{n-1} f(x)$$

iii) Distribution of rth order statistic $(X_{(r)})$, the general case :→

Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from a population having distribution function $F(\cdot)$.

Let $X_{(r)}$ denotes the rth order statistic, the p.d.f. of $X_{(r)}$ is given by,

$$f_{X_{(r)}}(x) = \lim_{h \downarrow 0} \frac{P\left[x - \frac{h}{2} < X_{(r)} < x + \frac{h}{2}\right]}{h}$$

$$= \lim_{h \downarrow 0} \frac{P\left[\begin{array}{l}(r-1) \text{ obs} < x - h/2, \text{ one obs. in } \left(x - \frac{h}{2}, x + \frac{h}{2}\right), \\ (n-r) \text{ obs.} > x + \frac{h}{2}\end{array}\right]}{h}$$

$$= \lim_{h \downarrow 0} \frac{n!}{(r-1)!\,(n-r)!} P\left[\text{A particular case}\right]$$

$$= \lim_{h \downarrow 0} \frac{n!}{(r-1)!\,(n-r)!} \cdot \frac{1}{h} P\left[\begin{array}{l} X_1, X_2, \ldots, X_{r-1} < x - \frac{h}{2}, \\ X_r \in \left(x - \frac{h}{2}, x + \frac{h}{2}\right), \\ X_{r+1}, X_{r+2}, \ldots, X_n > x + \frac{h}{2} \end{array}\right]$$

$$= \lim_{h \downarrow 0} \frac{n!}{(r-1)!\,(n-r)!} \left[F\left(x - \frac{h}{2}\right)\right]^{r} \frac{h \cdot f(x)}{h} \left[1 - F\left(x + \frac{h}{2}\right)\right]^{n-r}$$

$$= \frac{n!}{(r-1)!\,(n-r)!} \left\{F(x)\right\}^{r} \left\{1 - F(x)\right\}^{n-r} f(x)$$

- Alternative approach :→ Let $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ be the order statistics corresponding to the sample observation $(x_1, x_2, \ldots, x_n)$ having joint p.d.f.

$$f_\theta(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

if $g(y_1, y_2, \ldots, y_n)$ be the p.d.f. of $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ then

$$g(y_1, y_2, \ldots, y_n) = n!\, f(y_1, y_2, \ldots, y_n), \quad -\infty < y_1 < y_2 < \cdots < y_n < \infty.$$

Let, $F_r(y)$ be the distribution function of rth order statistic $X_{(r)}$.

Then, $F_r(y) = P\left[X_{(r)} \leq y\right]$

$= P$[at least r of the n th sample observation are $\leq y$]

$$= \sum_{t=r}^{n} \binom{n}{t} [F(y)]^{t} [1 - F(y)]^{n-t}, \text{ where } F \text{ is the d.f. of } X.$$

$$= 1 - \sum_{t=0}^{r-1} \binom{n}{t} [F(y)]^{t} [1 - F(y)]^{n-t}$$

$$= 1 - \frac{1}{\beta(n-r+1, r)} \int_0^{1-F(y)} z^{n-r} (1-z)^{r-1}\, dz$$

so that the pdf of $X_{(r)}$ is

$$f_r(y) = \frac{d}{dy} F_r(y) = [1-F(y)]^{n-r} [F(y)]^{r-1} \cdot f(y) \cdot \frac{1}{\beta(n-r+1, r)}$$

$$= \frac{n!}{(r-1)!\,(n-r)!} [F(y)]^{r-1} [1-F(y)]^{n-r} \cdot f(y)$$

Particular case:- Let, $r=1$, then the p.d.f. of minimum order statistic is, $f_1(y) = n[1-F(y)]^{n-1} f(y)$

Let, $r=n$, then the p.d.f. of maximum order statistic is,

$$f_n(y) = n\,[F(y)]^{n-1} f(y).$$

Example:-

1. Let $X \sim R(0,\theta)$ with pdf $f_\theta(x) = \begin{cases} \frac{1}{\theta} & ; 0<x<\theta \\ 0 & \text{ow} \end{cases}$

$\therefore$ PDf of $X_{(n)}$ is

$$f_n(y) = \begin{cases} n\left(\frac{y}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} & , 0<y<\theta \\ 0 & ; \text{ow} \end{cases}$$

and the pdf of $X_{(1)}$ is

$$f_1(y) = \begin{cases} n\left[1-\frac{y}{\theta}\right]^{n-1} \cdot \frac{1}{\theta} & ; 0<y<\theta \\ 0 & ; \text{ow} \end{cases}$$

2. Let $X \sim Exp(\theta, 1)$

$$\therefore f_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{if } x>\theta \\ 0 & \text{ow} \end{cases}$$

$\therefore$ PDF of $X_{(1)}$ is

$$f_1(y) = \begin{cases} n\left[1-\int_\theta^y e^{-(x-\theta)}dx\right]^{n-1} \cdot e^{-(y-\theta)} & \text{if } \theta<y<\infty \\ 0 & \text{ow} \end{cases}$$

Joint distribution of $X_{(1)}$ and $X_{(n)}$:-

The joint distribution function of $X_{(1)}$ and $X_{(n)}$ is given by,

$$F_{X_{(1)},X_{(n)}}(x,y) = P[X_{(1)} \le x, X_{(n)} \le y]$$

$$= P[X_{(n)} \le y] - P[X_{(1)} > x, X_{(n)} < y]$$

$$= P[X_1, X_2, \ldots, X_n \le y] - P[x < X_1, X_2, \ldots, X_n < y]$$

$$= [F(y)]^n - [F(y) - F(x)]^n .$$

$\therefore$ The joint pdf of $X_{(1)}$ and $X_{(n)}$ is given by,

$$f_{X_{(1)},X_{(n)}}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X_{(1)},X_{(n)}}(x,y)$$

$$= n(n-1)[F(y) - F(x)]^{n-2} f(x) f(y)$$

Joint pdf of $X_{(r)}$ and $X_{(s)}$, the general case :-

The joint PDF of $X_{(r)}$ and $X_{(s)}$ is given by,

$$f_{X_{(r)},X_{(s)}}(x,y) = \lim_{h\downarrow 0\, k\downarrow 0} \frac{1}{hk} P\left[x - \frac{h}{2} < X_{(r)} < x + \frac{h}{2},\ y - \frac{k}{2} < X_{(s)} < y + \frac{k}{2}\right]$$

if $r < s$.

$$= \lim_{\substack{h\downarrow 0\\ k\downarrow 0}} \frac{1}{hk} P\Big[(r-1) \text{ obs.} < x - \frac{h}{2}, \text{ one obs. in } \left(x - \frac{h}{2}, x + \frac{h}{2}\right), (s-r-1) \text{ obs. in } \left(x + \frac{h}{2}, y - \frac{k}{2}\right), \text{ one obs. in } \left(y - \frac{k}{2}, y + \frac{k}{2}\right), (n-s) \text{ obs} > y + \frac{k}{2}\Big]$$

$$= \lim_{\substack{h\downarrow 0\\ k\downarrow 0}} \frac{n!}{(r-1)!(n-s)!(s-r-1)!} P\left[\frac{\text{a particular case}}{hk}\right]$$

$$= \lim_{\substack{h\downarrow 0\\ k\downarrow 0}} \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \left\{F\left(x - \frac{h}{2}\right)\right\}^{r-1} \frac{h f(x)}{h} . \left\{F\left(y - \frac{k}{2}\right) - F\left(x + \frac{h}{2}\right)\right\}^{s-r-1} \frac{k . f(y)}{k} \left\{1 - F\left(y + \frac{k}{2}\right)\right\}^{n-s}$$

$$= \frac{n!}{(r-1)!(n-s)!(s-r-1)!} \{F(x)\}^{r-1} \{F(y) - F(x)\}^{s-r-1} \{1 - F(y)\}^{n-s} f(x) f(y) .$$

Sample Median & Sample Range :-

Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ denote the order statistics of a random sample $X_1, X_2, \ldots, X_n$ from a density $f(\cdot)$. The sample median is defined to be the middle order statistic if n is odd and and average of the middle two order statistics if n is even. The sample range is defined to be $X_{(n)} - X_{(1)}$, sample mid-range is defined to be $\{X_{(n)} + X_{(1)}\}/2$.

Example 1. Let $X_1, X_2, \ldots, X_n$ be iid RV's with common PDF

$$f(x) = \begin{cases} 1 & \text{if } 0<x<1 \\ 0 & \text{ow} \end{cases}$$

Find the distribution of sample Range.

ANS:- The joint PDF of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{X_{(1)}, X_{(n)}}(x,y) = n(n-1)\left(F(y)-F(x)\right)^{n-2}, \quad 0<x<y<1.$$

Now, $F(y) = \int_0^y 1\,dx = y.$

similarly, $F(x) = x$

$$\therefore f_{X_{(1)}, X_{(n)}}(x,y) = n(n-1)(y-x)^{n-2}, \quad 0<x<y<1.$$

Let us consider the following transformation,

$(X_{(1)}, X_{(n)}) \longrightarrow (X_{(1)}, R)$ such that

$$R = X_{(n)} - X_{(1)}$$

$$|J| = \left|\frac{\partial X_{(n)}}{\partial R}\right| = 1.$$

Here, $r = y - x$

$$\Rightarrow y = x + r$$

$$0<y<1$$

$$\Rightarrow 0<x<1-r, \quad 0<r<1.$$

The joint PDF of $X_{(1)}$ and R is given by,

$$f_{X_{(1)}, R}(x,r) = n(n-1)\, r^{n-2}, \quad \begin{array}{l} 0<x<1-r \\ 0<r<1 \end{array}$$

$\therefore$ PDF of R is given by,

$$f_R(r) = n(n-1)\, r^{n-2} \int_0^{1-r} dx, \quad 0<r<1$$

$$= n(n-1)\, r^{n-2}(1-r), \quad 0<r<1$$

Hence the answer.

Example 2. Find the distribution of the sample median in a random sample of size $n = 2K+1$, K be a positive integer, from a continuous distribution.

ANS:- Let $X_1, X_2, \ldots, X_{2K+1}$ be the sample drawn from a population having distribution function $F(\cdot)$.

The ordered sample is given by,

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(K-1)} \le X_{(K)} \le X_{(K+1)} \le \cdots \le X_{(2K+1)}$$

Clearly the random variable representing the sample median is given by $X_{(K)}$.

The pdf of $X_{(K)}$ is given by,

$$f_{X_{(K)}}(x) = \frac{(2K+1)!}{(K-1)!(K+1)!} \{F(x)\}^{K-1} \{1-F(x)\}^{K+1} f(x)$$

Example 3. Show that the order statistic $X_{(r)}, X_{(s)}, r < s$, the correlation coefficient is $\left[r(n-s+1)/s(n-r+1)\right]^{1/2}$, when the random sample is drawn from $U(0,1)$.

ANS:- A random sample of size n is drawn from $U(0,1)$.

$$\therefore f_{X_{(r)}}(x) = \frac{n!}{(r-1)!} x^{r-1} (1-x)^{n-r}, \quad 0 < x < 1$$

$$= \frac{x^{r-1}(1-x)^{n-r}}{B(r, n-r+1)}$$

$$\therefore E(X_{(r)}) = \frac{r}{n+1}.$$

$$Var(X_{(r)}) = \frac{r(n-r+1)}{(n+1)^2(n+2)}$$

Similarly, $Var(X_{(s)}) = \dfrac{s(n-s+1)}{(n+1)^2(n+2)}$

The joint PDF of $X_{(r)}$ and $X_{(s)}$ is given by,

$$f_{X_{(r)}, X_{(s)}}(x,y) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} x^{r-1} (y-x)^{s-r-1} (1-y)^{n-s}; \quad 0<x<y<1.$$

$$E\left[X_{(r)}, X_{(s)}\right] = \underbrace{\frac{n!}{(r-1)!\,(n-s)!\,(s-r-1)!}}_{(c)\ \text{cons.}} \times \iint_{0<x<y<1} xy\, x^{r-1} (y-x)^{s-r-1} (1-y)^{n-s}\, dx\, dy$$

$$= c \int_0^1 y(1-y)^{n-s} \left\{ \int_0^y x^r (y-x)^{s-r-1} dx \right\} dy$$

Let, $x = ty \Rightarrow dx = dt \cdot y$

$$= c \int_0^1 y(1-y)^{n-s} y^s \, dy \int_0^1 t^r (1-t)^{s-r-1} dt$$

$$= \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} B(r+1, s-r)\, B(s+2, n-s+1)$$

$$= \frac{r(s+1)}{(n+1)(n+2)}$$

$$\therefore \text{Cov}\left(X_{(r)}, X_{(s)}\right) = \frac{r(s+1)}{(n+1)(n+2)} - \frac{rs}{(n+1)^2}$$

$$= \frac{r(n-s+1)}{(n+1)^2(n+2)}$$

$\therefore$ Correlation coefficient between $X_{(r)}$ and $X_{(s)}$ is given by,

$$\rho = \frac{\frac{r(n-s+1)}{(n+1)^2(n+2)}}{\sqrt{\frac{s(n-s+1)}{(n+1)^2(n+2)}}\sqrt{\frac{r(n-r+1)}{(n+1)^2(n+2)}}} = \left[\frac{r(n-s+1)}{s(n-r+1)}\right]^{1/2}$$

Example 4. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics for a random sample of size n from the exponential distn. with pdf. $f(x) = \lambda e^{-\lambda x}$ if $x > 0$, where $\lambda > 0$. Show that $U_i = X_{(i)} - X_{(i-1)}$ for $i = 1, 2, \ldots, n$, $X_{(0)} = 0$, are independently distributed exponential variate and hence show that the expected value of the sample range is

$$\frac{1}{\lambda}\left[1 + \frac{1}{2} + \frac{1}{3} + \cdots\cdots + \frac{1}{n-1}\right]$$

Ans:- The joint PDF of $X_1, X_2, \ldots, X_n$ is given by

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, \quad x_i > 0$$

The joint Pdf of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is given by,

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_{(1)}, \ldots, x_{(n)}) = n!\, \lambda^n e^{-\lambda (x_{(1)} + x_{(2)} + \cdots + x_{(n)})},$$

$$0 \le x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \infty$$

Let us consider the following transformation,

$(X_{(1)}, X_{(2)}, \dots, X_{(n)}) \longrightarrow (U_1, U_2, \dots, U_n) \ni$

$$\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots -1 & 1 \end{pmatrix}}_{P} \begin{pmatrix} X_{(1)} \\ X_{(2)} \\ \vdots \\ X_{(n)} \end{pmatrix}$$

Jacobian of the transformation $= |P| = 1$.

$\therefore$ Joint pdf of $U_1, U_2, \dots, U_n$ is given by,

$$f_{U_1, U_2, \dots, U_n}(u_1, u_2, \dots, u_n) = n!\, \lambda^n e^{-\lambda(nu_1 + (n-1)u_2 + \cdots + u_n)}, \quad u_i > 0 \ \forall i$$

$$= n\lambda e^{-\lambda n u_1} (n-1)\lambda e^{-(n-1)\lambda u_2} \cdots \lambda e^{-\lambda u_n}$$

$$= f_{U_1}(u_1) \cdots f_{U_n}(u_n), \text{ as } U_i\text{'s are indep.}$$

$\therefore U_1 \sim \exp(n\lambda)$

$U_2 \sim \exp((n-1)\lambda)$

$\vdots$

$U_n \sim \exp(\lambda)$.

Now, $U_2 + U_3 + \cdots + U_n = X_{(n)} - X_{(1)}$

$= R$

$=$ Sample Range.

$$E(R) = E(U_2) + E(U_3) + \cdots + E(U_n)$$

$$= \frac{1}{\lambda} + \frac{1}{2\lambda} + \cdots + \frac{1}{(n-1)\lambda}$$

$$= \frac{1}{\lambda}\left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{(n-1)}\right]$$

Hence the proved.

Example 5. $X \sim R(0,1)$, find

i) the distn. of $X_{(r)}$.

ii) the m.g.f., mean $(E(X_{(r)}))$ and variance $(Var(X_{(r)}))$

ANS:- i) If $X \sim R(0,1)$, then the p.d.f. of X is given by,

$$f(x) = 1, \quad 0 < x < 1.$$

The distn. function of X is given by,

$$F(x) = \int_0^x dx = x$$

$\therefore$ the PDF of rth order statistic is given by,

$$g(x) = \frac{n!}{(r-1)!\,(n-r)!}\, x^{r-1} (1-x)^{n-r}, \quad 0 < x < 1.$$

$$\therefore X_{(r)} \sim B(r, n-r+1)$$

ii) Let $X_{(r)}$ be denoted as U, then

$$U \sim B(r-1, n-r)$$

$$M_U(t) = E(e^{Ut})$$

$$= E\left[1 + Ut + \frac{U^2t^2}{2!} + \cdots\cdots\right]$$

$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} E(U^r)$$

$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \cdot \mu_r' \quad \left[\because \mu_r' = r\text{th order raw moment about zero}\right]$$

$$E(U) = \frac{1}{B(r, n-r+1)} \int_0^1 u \cdot u^{r-1} (1-u)^{n-r}\, du, \quad 0 < x < 1$$

$$= \frac{B(r+1, n-r+1)}{B(r, n-r+1)}$$

$$= \frac{r}{n+1}.$$

$$E(U^2) = \frac{r(r+1)}{(n+1)(n+2)}$$

$$V(U) = \frac{r(r+1)}{(n+1)(n+2)} - \frac{r^2}{(n+1)^2} = \frac{r}{n+1}\left[\frac{r+1}{n+2} - \frac{r}{n+1}\right]$$

$$= \frac{r(n-r+1)}{(n+1)^2(n+2)}$$

Example 6. If X_1, X_2 be a random sample of size 2 drawn from a population having p.d.f. $f(x) = \lambda e^{-\lambda x}, x>0, \lambda>0$. Then find the distn of the sample range. Is the distn independent from the sampling distribution of simple AM?

Ans:- x_1, x_2 be a random sample drawn from a popln with pdf
$$f(x) = \lambda e^{-\lambda x}, \lambda>0, x>0.$$
Let us consider the following transformation
$$(X_1, X_2) \to (X_{(1)}, X_{(2)}), \text{ where } X_{(i)} = i\text{th order statistic}.$$
$\therefore$ Joint distn of $X_{(1)}, X_{(2)}$ is given by,
$$f_{X_{(1)}, X_{(2)}}(x_1, x_2) = 2\lambda^2 e^{-\lambda(x_1+x_2)}$$
Let us define a variable,
$$U_i = X_{(i)} - X_{(i-1)} \quad \forall i = 1, 2.$$
$$U_1 = X_{(1)} \quad [\text{Assuming } X_{(0)} = 0]$$
$$U_2 = X_{(2)} - X_{(1)}$$
$$\therefore X_{(2)} = u_1 + u_2$$
$$\therefore |J| = \left| J\left(\frac{X_{(1)}, X_{(2)}}{u_1, u_2}\right) \right| = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$
$\therefore$ Joint pdf of U_1, U_2 is given by,
$$f_{U_1, U_2}(u_1, u_2) = 2\lambda^2 e^{-\lambda(2u_1+u_2)}$$
$$= 2\lambda . e^{-2\lambda u_1} . \lambda e^{-\lambda u_2}, (u_1, u_2) > 0$$
$\therefore U_1, U_2$ are independently distributed.

$\therefore$ Sample range $(R) = X_{(2)} - X_{(1)}$
$$= u_2.$$
$$\therefore E(R) = E(u_2) = \lambda \int_0^\infty u e^{-\lambda u_2} du$$
$$= \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$
$\therefore$ PDF of sample range (R) is $f_R(R) = \lambda e^{-\lambda R}, R>0$

Now, simple AM $= \dfrac{X_{(1)} + X_{(2)}}{2} = \dfrac{2u_1 + u_2}{2} = u_1 + \frac{1}{2}u_2 = Z$, say.

$\therefore$ Joint PDF of (Z, u_2) is given by,
$$f_{Z, u_2}(z, u_2) = 2\lambda^2 e^{-2\lambda z}, z>0$$
$\therefore$ PDF of Z is, $f_Z(z) = 2\lambda^2 e^{-2\lambda z}, z>0$

So, the distn. of sample range & simple AM are different.