→ Gamma distribution naturally occurs in the processes where the waiting times between events are relevant

→ It is used to predict the wait time until future events occur

→ Gamma distribution predicts the wait time until the kth (k^{th}) event occurs.

- Gamma Function:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx \quad , \quad \alpha > 0$$

$\Gamma(n) = (n-1)!$ $\qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$\Gamma(1) = 1$

$\Gamma(n+1) = n\,\Gamma(n)$

- ~~$f(x, \alpha,$~~ Gamma distribution :

$$f(x, \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & , \ x > 0 \\ 0 & , \ \text{elsewhere} \end{cases}$$

$\alpha > 0 \ , \ \beta > 0$

- Exponential distribution: ($\alpha = 1$ in Gamma)

$$f(x, \beta) = \begin{cases} \dfrac{1}{\beta} e^{-\frac{x}{\beta}} & , \ x > 0 \\ 0 & , \ \text{elsewhere} \end{cases}$$

$\beta > 0$.

Gamma distribution	Exponential distribution
$\mu = \alpha\beta$	$\mu = \beta$
$\sigma^2 = \alpha\beta^2$	$\sigma^2 = \beta^2$

Page No.
Date / 20

- Integration by parts:

$$\int u\,v\,dx = u\int v\,dx - \int \left(\frac{du}{dx}\right)\cdot\left(\int v\,dx\right)dx$$

- Use L Hospital rule.

- MGF of Gamma distribution:

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx}\frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}\,dx$$

$$\boxed{M_X(t) = (1-t\beta)^{-\alpha}}$$

Mean $\Rightarrow E(X) = \mu = M'_X(0) = \alpha\beta$

Variance $\Rightarrow Var(X) = E(X^2) - (E(X))^2 =$

$= M''_X(0) - (M'_X(0))^2$

$= \alpha\beta^2$

- MGF of Exponential dist:- $\alpha = 1$

$$M_X(t) = (1-t\beta)^{-1}$$

$\mu = \beta$, $\sigma^2 = \beta^2$

Page No.
Date / 20

1. Unbiasedness.

$$E(T) = \theta$$

{Estimator $T(x_1, x_2, \dots, x_n) \Rightarrow$ an unbiased estimator of θ}

$E(T) > \theta$ +vely biased estimator
$E(T) < \theta$ -vely biased estimator.

Bias $\Rightarrow$ $b(\theta) = E(T) - \theta$

$\boxed{E(T(x)) = \sum_{x=0}^{\infty} T(x) f(x)}$

For Binomial $\Rightarrow$

$T = \sum_{i=1}^{n} x_i \sim B(n, \theta)$

For Normal $\Rightarrow$

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

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$$e^{-x} = 1 - \frac{x}{1!} + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots$$

$Var(T) = E(T^2) - [E(T)]^2$

2. Consistency:

(i) $E(T_n) \to \theta$ as $n \to \infty$.
(ii) $Var(T_n) \to 0$ as $n \to \infty$

$\boxed{T' = \left(\frac{n-a}{n-b}\right) T_n \quad n \to \infty, T_n \to \theta}$

T_n is a consistent estimator of θ.

Invariance property of consistent estimator;

$T_n \to \theta$ (A consistent estimator)

$\psi(\theta) \Rightarrow$ continuous function of θ.

$\psi(T_n) \longrightarrow \psi(\theta)$

$\hookrightarrow$ A consistent estimator

Sample mean is a consistent estimator of population mean

3. Efficiency

$V(T_1) < V(T_2)$ for all n

T_1 is more efficient than T_2

$V(\bar{x}) = \frac{\sigma^2}{n}$	$T_1 \Rightarrow$ Most efficient estimator of variance V_1
$V(m_d) = \frac{\pi}{2}\frac{\sigma^2}{n}$ (median)	$T_2 \Rightarrow$ Other estimator of variance V_2
$= 1.57 \frac{\sigma^2}{n}$	Efficiency of T_1 of $T_2 \Rightarrow$
$V(\bar{x}) < V(m_d)$ → more efficient	$E = \frac{V_1}{V_2}$ → can not exceeds unity

4. Sufficiency

Factorization theorem (Neymann) :-

$$L = g_\theta[t(x)] \cdot h(x)$$

Invariance property of Sufficient estimator $\Rightarrow$ $T \rightarrow$ sufficient estimator for θ

$\psi(T) \rightarrow$ " " $\rightarrow \psi(\theta)$

One to One

Fisher - Neymann criteria :

$$L = \prod_{i=1}^{n} f(x_i, \theta) = \underbrace{g_1(t_1, \theta)}_{\text{pdf of statistics } t_1} \cdot \underbrace{K(x_1, x_2 \cdots x_n)}_{\text{function of sample observation}}$$

pdf

function of sample observation = independent of θ

Page No.
Date / 20

• Central Limit Theorem:

Form 1: $E(X_i) = \mu_i$ $\quad Var(X_i) = \sigma_i^2$

$S_n = X_1 + X_2 + \cdots + X_n$

Then,
$S_n \Rightarrow$ follows $\Rightarrow$ Normal distribution
$\Downarrow$
with mean $\mu = \sum_{i=1}^{n} \mu_i$ &

Variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$

as $n \to \infty$

Form 2: $E(X_i) = \mu$ $\quad Var(X_i) = \sigma^2$

$S_n = X_1 + X_2 + \cdots + X_n$

Then,
$S_n \Rightarrow$ follows $\Rightarrow$ Normal distribution
$\Downarrow$
with mean $\Rightarrow n\mu$ &

Variance $\Rightarrow n\sigma^2$

as $n \to \infty$

As $n \to \infty$, $X \sim N(\mu, \sigma^2)$

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

Maximum Likelihood Estimator

Likelihood function ⇒

$$L = f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

maximize the likelihood function →

$$\boxed{\frac{\partial \log L}{\partial \theta} = 0} \qquad \boxed{\frac{\partial L}{\partial \theta} = 0 \quad \frac{\partial^2 L}{\partial \theta^2} < 0}$$

$L > 0$,
Log L is nondecreasing funcⁿ of L

So, L & Log L attain their extreme value at the same point

→ If MLE exists ⇒ Most efficient.

→ Invariance property of MLE ⇒ T is MLE of θ

$\psi(T)$ is MLE of $\psi(\theta)$

One to One

→ MLE's ⇒ are always consistent but need not be unbiased

Student's t test for single mean.

1. $$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

s - If standard deviation given in ques

S - If not given in ques.

population mean $\Rightarrow \mu_0$
sample mean $\Rightarrow \bar{x}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

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

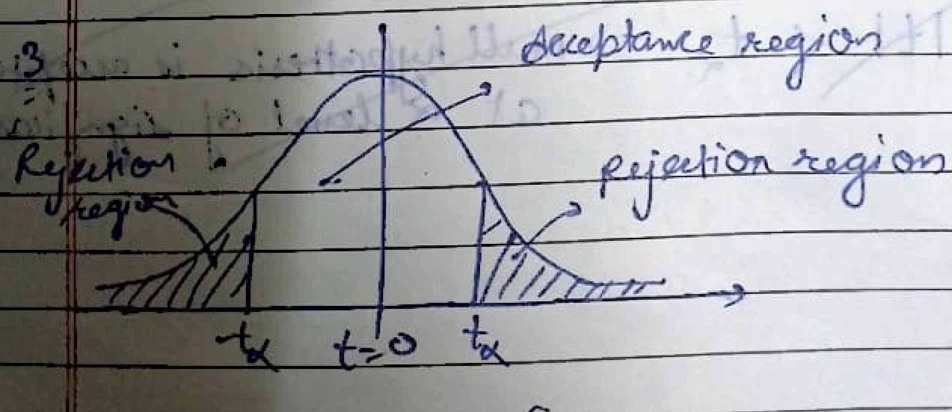
$$n s^2 = (n-1) S^2$$

Assume
- Normal
- Random sample
- σ is unknown

Degree of freedom $= n-1$

2. Calculated $|t| >$ tabulated t
(Null hypothesis is rejected)

Calculated $|t| <$ tabulated t
{Null hypothesis may be accepted at the level of significance adopted}

3. 

4. Confidence limits (95% if)

$$\bar{x} \pm t_{0.05} \frac{S}{\sqrt{n}}$$

t test difference of mean.

$x_i\ (i = 1, 2, 3, \ldots n_1) \Rightarrow \mu_x$

5. $\sigma_x^2 = \sigma_y^2 = \sigma^2$ $\quad y_i\ (i = 1, 2, 3, \ldots n_2) \Rightarrow \mu_y$

6. $$t = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

7. $$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right]$$

$$= \frac{1}{n_1 + n_2 - 2}\left[n_1 s_x^2 + n_2 s_y^2\right]$$

8. degree of freedom $= n_1 + n_2 - 2$

9. S^2 is an unbiased estimator of the common population variance σ^2

10. ~~tab. $|t| > t_\alpha \Rightarrow$ Null hypothesis is rejected~~

~~tab $|t| < t_\alpha \Rightarrow$ Null hypothesis is accepted at α level of significance~~

$O_i \Rightarrow$ Set of observed (experimental) frequencies

$E_i \Rightarrow$ Corresponding set of expected (theoretical or hypothetical) frequencies.

- Chi-square test for goodness of fit:

Karl Pearson's

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] \qquad \left(\sum_{i=1}^{n} O_i = \sum_{i=1}^{n} E_i\right)$$

Degree of freedom $= n-1$

Does not make any assumptions regarding parent population

Distribution free test / Non parametric

- F-test

$$F = \frac{S_{x^2}}{S_{y^2}}$$

(Numerator is always greater)

$$S_{x^2} = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$$

$$S_{y^2} = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(y_i - \bar{y})^2$$

$F(n_1 - 1, n_2 - 1)$

From (graph)

The t-test is used to compare the means of two groups & determine if they are significantly different,

while the F-test is used to compare variances of two or more groups and assess if they are significantly different.

No of degree of freedom = No of observations − No of independent constraints

- Z - test.

$n \geq 30 \Rightarrow z$-test
$n < 30 \Rightarrow t$ test

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad [\sigma \text{ is known}]$$

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim N(0,1) \quad [\sigma \text{ is unknown}]$$

z test for difference of mean.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

	1%	5%	10%
Two tailed	2.58	1.96	1.645
Right tailed	2.33	1.645	1.28
Left tailed	-2.33	-1.645	-1.28