

Recall  $\frac{d}{d\theta} \log L(\underline{x} | \theta) = 0$

$$-n + \frac{\sum x_i}{\lambda} = 0$$

Recall  $\sum x_i = n\bar{x}$

$$-n + \frac{n\bar{x}}{\lambda} = 0$$

$$-n = -\frac{n\bar{x}}{\lambda}$$

$$\lambda = \frac{\bar{x}}{n}$$

$$\lambda = \bar{x} = \hat{\lambda} = \bar{x} \text{ (sample mean)}$$

Prove that  $\bar{x}$  is the MLE for  $\lambda$

Recall

$$\frac{d^2}{d\theta^2} \log L(\underline{x} | \theta) < 0 \text{ ic its negative}$$

$$\frac{d^2}{d\lambda^2} \log L(\underline{x} | \lambda) = \frac{d}{d\lambda} \left( -n + \frac{\sum x_i}{\lambda} \right)$$

$$\frac{d}{d\lambda} \left( -n + \frac{n\bar{x}}{\lambda} \right)$$

$$\frac{d}{d\lambda} (-n + n\bar{x}\lambda^{-1})$$

$$= 0 + (-1)(1)n\bar{x}\lambda^{-2} = -n\bar{x}\lambda^{-2} = -\frac{n\bar{x}}{\lambda^2} < 0$$

hence  $\bar{x}$  is the MLE for  $\lambda$

Obtain the MLE of parameter  $f(x; \alpha) = \begin{cases} \alpha e^{-\alpha x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$  in an exponential population given by

Solution

Let  $x_1, x_2, \dots, x_n$  be a random sample from the exponential population.

The likelihood function  $L(\underline{x}, \alpha) = \prod_{i=1}^n \alpha e^{-\alpha x_i}$

$$L(\underline{x}; \alpha) = \alpha^n e^{-\alpha \sum_{i=1}^n x_i}$$

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

$$L(\underline{x}; \alpha) = \alpha^n e^{-\alpha n \bar{x}}$$

Get the log likelihood function

$$\begin{aligned} \log L(\underline{x}; \alpha) &= \log (\alpha^n e^{-\alpha n \bar{x}}) \\ &= n \log \alpha - \alpha n \bar{x} \end{aligned}$$

We get the first derivative of the log likelihood function = 0

$$\frac{d}{d\alpha} \log L(\underline{x}; \alpha) = \frac{d}{d\alpha} (n \log \alpha - \alpha n \bar{x})$$

$$\frac{d}{d\alpha} \log f(x) = \frac{f(x)'}{f(x)}$$

$$= \frac{n(1)}{\alpha} - n \bar{x} = 0$$

$$\frac{n}{\alpha} = n \bar{x}$$

$$\hat{\alpha} = \bar{x} \quad \alpha = \frac{1}{\bar{x}} = \frac{1}{\text{sample mean}}$$

Prove / verify that  $\hat{\alpha} = \frac{1}{\bar{x}}$  (MLE)

$$\frac{d^2}{d\alpha^2} \log L(\underline{x}, \alpha) = \frac{d}{d\alpha} \left( \frac{n}{\alpha} - n\bar{x} \right)$$

$$= \frac{d}{d\alpha} n\alpha^{-1} - n\bar{x}$$

$$= (-1)(1)n\alpha^{-2} = -n\alpha^{-2} = \frac{-n}{\alpha^2} \quad (\text{negative})$$

$< 0$  hence  $\hat{\alpha} = \frac{1}{\bar{x}}$  is the MLE

Exercise,

$X$  follows a normal distribution  $N(\mu, \sigma^2)$ . Let  $X_1, X_2$  upto  $X_n$  be a random sample from this population. find the MLEs of

- i)  $\mu$  when  $\sigma^2$  is known
- ii) Get the MLE of  $\sigma^2$  when  $\mu$  is known
- iii) Get the " of  $\mu$  &  $\sigma^2$  when both are unknown.

## Properties of estimators

A good estimator should

- i) unbiasedness
- ii) consistency
- iii) sufficiency
- iv) mean square error
- v) efficiency.

passes the following properties

### 1. Unbiasedness

An estimator Suppose  $X_1, X_2, \dots, X_n$  is a random sample of size  $n$  from a distribution having a density function  $f(x, \theta)$  where  $\theta$  is a known parameter.

Again assume that  $\theta$  is an element of sample space  $\Omega$ , which is a subset of the real life then an estimator  $\hat{\theta}$  is said to be unbiased estimator of  $\theta$  if the expected value of  $\hat{\theta} = \theta$  i.e.  $E(\hat{\theta}) = \theta$  for all  $\theta$  in sample space  $\Omega$ .

The difference of the expected value of theta but and  $\theta$  is called biasness of theta  $E(\hat{\theta}) - \theta = b(\theta)$ .

This means that if the sampling mean of sampling distribution of estimator  $\hat{\theta}$  is  $\theta$  then  $\hat{\theta}$  is an unbiased estimator of  $\theta$ .

### Example

Suppose  $X_1, X_2, \dots, X_n$  is a random sample from a uniform population with density function  $f(x) = \begin{cases} 1 & \lambda - \frac{1}{2} < x < \lambda + \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$

Prove that  $\bar{X}$  is an unbiased estimator of  $\lambda$

Solution

$\bar{X}$  is unbiased estimator of  $\lambda$

$$E(\hat{\theta}) = \theta$$

$$E(X) = E(\bar{X}) = \lambda$$

$$\text{Recall } E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

$$E(\bar{X}) = E(X) = \int x dx$$

$$= \int_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}} x \cdot dx = \frac{x^2}{2} \Big|_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}}$$

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

$$(\lambda + \gamma_2)(\lambda + \frac{1}{2}) = \lambda^2 + \frac{1}{2}\lambda + \gamma_2\lambda + \frac{1}{4}$$
$$= \lambda^2 + \lambda + \frac{1}{4}$$

$$(\lambda - \gamma_2)(\lambda - \frac{1}{2}) = \lambda^2 - \frac{1}{2}\lambda - \frac{1}{2}\lambda + \frac{1}{4}$$
$$= \lambda^2 - \lambda + \frac{1}{4}$$

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

$\therefore \bar{X}$  is unbiased estimator of  $\lambda$

Suppose that  $x_1, x_2, \dots, x_n$  is random sample from population & with pdf given by  $f(x) = \begin{cases} 1/\theta e^{-x/\theta} & \text{for } x, \theta > 0 \\ 0, \text{ elsewhere} \end{cases}$

Show that  $\bar{x}$  is unbiased estimator of  $\theta$

Solution

$$E(\bar{x}) = E(X) = \theta$$

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

$$= \int_0^{\infty} x \cdot \frac{1}{\theta} e^{-x/\theta} dx \quad \text{---} \theta$$

Method 1

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

$$\text{let } Y = \frac{X}{\theta} = X = \theta Y \text{ or } Y\theta$$

$$\frac{dy}{dx} = \frac{1}{\theta} \Rightarrow \theta dy = dX$$

$$= \frac{1}{\theta} \int_0^{\infty} \cancel{x} e^{-\cancel{x}/\theta} \theta dy \quad \left( \frac{1}{\theta} \int_0^{\infty} Y \theta e^{-Y} \theta dy \right)$$

$$\underbrace{\frac{\theta^2}{\theta} \int_0^{\infty} Y e^{-Y} dy}_{\Gamma_2} = (n-1)! = (2-1)! = 1! = 1$$

$$= \frac{\theta^2}{\theta} = \theta$$

Exercise.

- 1) Repeat above qsn using integration by parts
- 2) Suppose  $x_1, x_2, \dots, x_n$  is random sample of size  $n$  from distribution with mean  $\mu$  and variance  $\sigma^2$ .  
Show that  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  is unbiased estimator of  $\sigma^2$   
and  $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  is a biased estimator of  $\sigma^2$ .

Estimable function

- Estimators (properties)
- i) Unbiasedness  $E(\hat{\theta}) = \theta$
  - ii) Consistent  $E(\hat{\theta}) = \theta$  or  $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$
  - iii) Sufficiency  $\lim \text{var}(\hat{\theta}) = 0$
  - iv) Efficiency - MSE

Ex 1

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

$$E(\bar{x}) = \mu$$

$$(\bar{x}) = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$= \underbrace{E(x_1 + x_2 + x_3 + \dots + x_n)}_n$$

$$= 1/n (E(x_1) + E(x_2) + \dots + E(x_n))$$

$$\text{recall } E(x) = \mu$$

$$1/n (\mu_1 + \mu_2 + \dots + \mu_n) = \frac{1}{n} (n\mu)$$

$$E(\bar{x}) = \mu$$

$$2. \lim_{n \rightarrow \infty} \text{var}(\bar{x}) = 0$$

$$\text{var}(\hat{\theta}) = \text{var}(\bar{x}) = \text{var}(x_1 + x_2 + \dots + x_n)$$

$$\text{var}(x_1 + \dots + x_n) = \frac{1}{n^2} \text{var}(x)$$

$$\text{var}(x_1 + x_2 + \dots + x_n) = \text{var}(x_1) + \text{var}(x_2) + \dots + \text{var}(x_n)$$

$$= \frac{1}{n^2} [\text{var}(x_1) + \text{var}(x_2) + \dots + \text{var}(x_n)]$$

$$\cdot \frac{1}{n^2} [\sigma^2 + \sigma^2 + \dots + \sigma^2]$$

recall

$$\text{Var } Y = \delta^2$$
$$\frac{1}{n} \sum (n \delta^2) = \frac{\delta^2}{n}$$

$$\lim_{n \rightarrow \infty} \frac{\delta^2}{n} = \frac{\delta^2}{\infty} = 0$$

- 4) Let  $X_1, X_2, \dots, X_n$  be a random sample of size  $n$  from a uniform population given by
- $$f(x) = \begin{cases} 1, & \lambda - \frac{1}{2} < x < \lambda + \frac{1}{2} \\ 0, & \text{elsewhere} \end{cases}$$

Show that  $\bar{x}$  is a consistent estimator of  $\lambda$ .

Solution

$$E(\hat{\theta}) = \theta$$

This means we need to get mean of this distribution

$$\text{mean} = E(X)$$

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

$$E(X) = \int_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}} x \cdot dx = \frac{x^2}{2} \Big|_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}}$$

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

$$\frac{(\lambda + \frac{1}{2})(\lambda + \frac{1}{2})}{2} - \frac{(\lambda - \frac{1}{2})(\lambda - \frac{1}{2})}{2}$$

$$\frac{2\lambda}{2} = \lambda$$

$$E(\bar{X}) = E(X) = \lambda$$

Condition 2  
 $\text{Var}(\bar{X}) = 0$   
 $n \rightarrow \infty$

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

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

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

$$= \int_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}} x^2 dx$$

$$= \frac{x^3}{3} \Big|_{\lambda - \frac{1}{2}}^{\lambda + \frac{1}{2}}$$

$$\frac{(\lambda + \frac{1}{2})^3}{3} - \frac{(\lambda - \frac{1}{2})^3}{3}$$

$$\frac{(\lambda + \frac{1}{2})^3}{3} (\lambda + \frac{1}{2}) - \frac{(\lambda - \frac{1}{2})^3}{3} (\lambda - \frac{1}{2})$$

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

$$\frac{\lambda^3 + \frac{1}{2}\lambda^2 + \lambda^2 + \frac{1}{2}\lambda + \frac{1}{4}\lambda + \frac{1}{8}}{3} - \frac{\lambda^3 - \frac{1}{2}\lambda^2 - \lambda^2 + \frac{1}{2}\lambda + \frac{1}{4}\lambda - \frac{1}{8}}{3}$$

$$\frac{\lambda^3 + \frac{3}{2}\lambda^2 + \frac{3}{4}\lambda + \frac{1}{8}}{3} - \frac{\lambda^3 - \frac{3}{2}\lambda^2 + \frac{3}{4}\lambda - \frac{1}{8}}{3}$$

$$\frac{\frac{3\lambda^2 + \frac{1}{4}}{3}}{3} = \lambda^2 + \frac{1}{12}$$

$$E(X^2) = \lambda^2 + \frac{1}{12}$$

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

$$= \lambda^2 + \frac{1}{12} - \lambda^2$$

$$= \frac{1}{12}$$

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$$

$$\lim_{n \rightarrow \infty} \frac{1}{12} = 0.083$$

$\therefore \bar{X}$  is a consistent estimator of  $\lambda$

### 3. Sufficiency.

An estimator  $T = t(X_1, X_2, \dots, X_n)$  of a parameter  $\theta$  based on a random sample  $X_1, X_2, \dots, X_n$  from a population with R.d.f  $f(x, \theta)$  is sufficient for  $\theta$  if the conditional distribution of  $X_1, X_2, \dots, X_n$  given  $t$  is independent of  $\theta$ , ie  $P(X_1=x_1, X_2=x_2, \dots, X_n=x_n | T)$  is independent of  $\theta$ .

An estimator  $T = t(X_1, X_2, \dots, X_n)$  is a sufficient estimator of parameter  $\theta$  if it contains all the information in a random sample about the parameter.

#### Factorisation criterion / Neyman theorem

This is a method of getting the necessary and sufficient condition for an estimator to be sufficient. It states that  $T = t(X_1, X_2, \dots, X_n)$  is a sufficient estimator for parameter  $\theta$  if and only if the joint density function (Likelihood function) - L of sample values  $X_1, X_2, \dots, X_n$  can be written in form of

$$L = L(\theta) = L(X, \theta) = L(X_1, X_2, \dots, X_n, \theta) = g(T, \theta) h(x)$$

where i)  $g(T, \theta)$  depends on  $\theta$  and  $x$  only through the value function  $T = t(X)$

ii)  $h(x)$  does not depend on  $\theta$

iii)  $g(t, \theta)$  and  $h(x)$  are positive functions ie  $> 0$

NB A function is independent of parameter  $\theta$  if and only if (iff) it does not contain the parameter  $\theta$  in its pdf as well as in its domain.

Eg

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

$$f(x) = \frac{1}{2\theta} \quad \text{for } 0 < x < 2$$

2<sup>nd</sup> function is independent of parameter  $\theta$

Steps of finding a sufficient estimator.

Step 1

Get the likelihood function ie  $L = L(X, \theta)$

$$L = \prod_{i=1}^n f(x_i, \theta)$$

Step 2

Check whether the likelihood function satisfies the factorisation criteria.

$$\text{ie } \Rightarrow L = g(T, \theta) h(x)$$

i)  $h(x)$  is independent of parameter  $\theta$

$$\text{ii) } h(x) \neq g(T, \theta) > 0$$

Step 1 If so  $T = t(x_1, x_2, \dots, x_n)$  is a sufficient estimator of the parameter  $\theta$

Example 1

Let  $x_1, x_2, \dots, x_n$  be a random sample of size  $n$  from exponential distribution defined by  $f(x) = \frac{1}{\theta} e^{-x/\theta}$  for  $x, \theta > 0$ . Find the sufficient estimator of parameter  $\theta$

Solution

Step 1

$$L = L(X, \theta)$$

$$L = \prod_{i=1}^n f(x_i, \theta)$$

$$= \frac{1}{\theta^n} e^{-\sum x_i / \theta} \times \frac{1}{\theta} e^{-x_1 / \theta} \times \dots \times \frac{1}{\theta} e^{-x_n / \theta}$$

$$\left(\frac{1}{\theta}\right)^n \left(e^{-\frac{\sum_{i=1}^n x_i}{\theta}}\right) g(T, \theta)$$

$$L(\theta) = \left(\frac{1}{\theta}\right)^n e^{-\frac{\sum_{i=1}^n x_i}{\theta}} \times Q_{hx}$$

$\sum_{i=1}^n x_i$  is a sufficient estimator for  $\theta$

Note : Sufficient estimator always takes the form of  $\sum_{i=1}^n x_i$  or  $\prod_{i=1}^n x_i$

### Example 2

Let  $x_1, x_2$  upto  $x_n$  be a random sample of size  $n$  from a population with pdf  $f(x_i; \theta) = \theta x_i^{\theta-1}$  for  $0 < x < 1$ ,  $\theta > 0$ . Find the sufficient estimator for parameter  $\theta$

### Solution 1

Get the likelihood function

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

$$= \theta x_1^{\theta-1} \times \theta x_2^{\theta-1} \times \dots \times \theta x_n^{\theta-1}$$

$$= \theta^n \prod_{i=1}^n x_i^{\theta-1}$$

$$\theta^n \prod_{i=1}^n x_i^{\theta-1} = \theta \left( \prod_{i=1}^n x_i \right)^\theta \times \prod_{i=1}^n (x_i)^{\theta-1}$$

$$L(\theta) = \underbrace{\theta^n \left( \prod_{i=1}^n x_i \right)^\theta}_{g(T, \theta)} \times \underbrace{\frac{1}{\prod_{i=1}^n x_i}}_{h(x)}$$

$$\text{where } h(x) = \frac{1}{\pi^n} \prod_{i=1}^n x_i$$

$$g(T, \theta) = \theta^n \left( \prod_{i=1}^n x_i \right)$$

$h(x)$  and  $g(T, \theta) > 0$

The sufficient estimator for  $\theta = \frac{\prod_{i=1}^n x_i}{\pi^n}$

### Example / exercise

Let  $X_1, X_2, \dots, X_n$  be a random sample from a binomial distribution with parameter  $\theta$

$$f_X = \begin{cases} \theta^x (1-\theta)^{n-x} & \text{for } x=0, 1 \\ 0 & \text{elsewhere} \end{cases}$$

$$T = \sum_{i=1}^n X_i$$

2. Confidence for mean  $\mu$  for t-distribution (ie when  $< \infty$ )  
 i) For one tailed test.

$$\bar{X} \pm t(\alpha, n-1) \times \frac{s}{\sqrt{n}}$$

where  $\bar{X}$  = sample mean.

$t(\alpha, n-1)$  = tabulated t value at  $\alpha$  level of significance and  $(n-1)$  degrees of freedom

$s$  = sample standard deviation

$n$  = sample size

$$= \left[ \bar{X} - t(\alpha, n-1) \times \frac{s}{\sqrt{n}} < \mu < \bar{X} + t(\alpha, n-1) \times \frac{s}{\sqrt{n}} \right]$$

ii) The 2 tailed test

$$\bar{X} \pm t(\alpha/2, n-1) \times \frac{s}{\sqrt{n}}$$

$$\left[ \bar{X} - t(\alpha/2, n-1) \times \frac{s}{\sqrt{n}} < \mu < \bar{X} + t(\alpha/2, n-1) \times \frac{s}{\sqrt{n}} \right] = (1-\alpha)$$

$\alpha$  = level of significance

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

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

Example

The table below shows a sample of 10 patients in a heart study.

Sample size  
 $n=10$

Sample mean  
121.2

Std

11.1

Construct a 95% confidence interval for the true systolic

i) For one tailed test

$$\bar{X} \pm t(\alpha, n-1) \times \frac{s}{\sqrt{n}}$$
$$121.2 \pm 1.833 \times \frac{11.1}{\sqrt{10}}$$
$$121.2 \pm 6.434064996$$

$$114.765935 < u < 127.67$$

We are 95% sure that the true systolic blood pressure lies between 114.7 and 127.6

ii) For 2 tailed test

$$(t_{\alpha/2}, n-1)$$

$$(t_{25\%}, 9)$$

$$= 2.262$$

$$121.2 \pm 2.262 \times \frac{11.1}{\sqrt{10}}$$

$$121.2 \pm 7.939909995$$

$$129.13991 < u < 113.26009$$

We are 97.5% sure that the true systolic blood pressure lies between 113.26 and 129.13991

iii) Test the hypothesis that

$$H_0 : \text{Systolic blood pressure} = 130$$

$$H_1 : \text{Systolic blood pressure} < 130$$

$H_1$  ✓ The systolic blood pressure is  $< 130$

H/w Go repeat when  $n$  is (Sample size) 3439 and compare the the values -

CIs

NOTE when you increase sample size error reduces  
" " decrease " " increases

Confidence Interval for difference in means.  
(samples obtained from some normal population)

The  $(1-\alpha) 100\%$  CI for difference in means obtained from the same population is given by

$$(\bar{X} - \bar{Y}) \pm z_{\frac{\alpha}{2}} \delta \sqrt{\frac{m+n}{mn}}$$

which is equivalent to

$$(\bar{X} - \bar{Y}) - z_{\frac{\alpha}{2}} \delta \sqrt{\frac{m+n}{mn}} < u_1, -u_2 < (\bar{X} - \bar{Y}) + z_{\frac{\alpha}{2}} \delta \sqrt{\frac{m+n}{mn}}$$

where  $\bar{X}$  = sample mean for  $x_1, x_2, \dots, x_m$

$\bar{Y}$  = sample mean for  $y_1, y_2, \dots, y_n$

$m$  = sample size for  $X$

$n$  = sample size for  $Y$

$\delta$  = common standard deviation

$z_{\frac{\alpha}{2}}$  = Z-score at  $\frac{\alpha}{2}$   $\alpha$  = levels of significance

Example

In 46 test runs the fuel consumption of an experimental engine had a std at 2 gallons  
D) Construct a 99% CI for

A sample of 20 light bulbs of one kind lasted on average of 530 hours of continuous use. Another sample of 30 bulbs of different kind lasted on avg 510 hours of continuous use. The 2 samples of light bulbs have a common standard deviation of 25. Construct a 99% CI for the difference of the average hours of use for the 2 kinds of light bulbs

Solution

$$(\bar{x} - \bar{y}) - z_{\alpha/2} \cdot \frac{\delta}{\sqrt{m+n}} < u_1 - u_2 < (\bar{x} - \bar{y}) + z_{\alpha/2} \cdot \frac{\delta}{\sqrt{m+n}}$$

$$(530 - 510) - 2.5758 \times 25 \sqrt{\frac{20+30}{20 \times 30}} < (530 - 510) + 2.5758 \times 25 \sqrt{\frac{20+30}{20 \times 30}}$$

$$20 - 2.5758 \times 25 \sqrt{\frac{50}{600}} < u_1 - u_2 < 20 + 2.5758 \times 25 \sqrt{\frac{50}{600}}$$

$$20 - 18.58923529 < u_1 - u_2 < 38.58923529$$

$$1.41076$$

We are 99.5% sure that the true difference between  $u_1$  and  $u_2$  lies between that range

#### 4) CI for Proportions

If sampling is done from an infinite population or is done with replacement from a finite population the CI for a proportion is given by

$$P \pm Z_{\alpha/2} \sqrt{Pq/n}$$

where  $P$  is the proportion of success in the sample

$$q = 1 - p$$

$n$  - Sample size.

#### Example 1

The human resource director of oracle operation wanted to know what proportion of all persons who have ever been interviewed for a job within corporation had been hired. He settled at 95% for CI. Random sample of 500 interview records revealed that 76 persons in the sample had been hired. Construct the interval

Ans

$$(13, 19) \text{ persons}$$

$$(0.120, 0.180346) \text{ proportion}$$

Example 2.

Out of 20,000 customer ledger account a sample of 600 accounts was taken to a test. The test was for accuracy of posting and balancing. Where 45 mistakes were recorded. Assign limits within which the number of defective cases will lie at  $\alpha = 5\%$ .

$$P =$$

5) CI for variance

The MLE for  $s^2$  has a chi-square ( $\chi^2$ ) distribution with  $n-1$  degrees of freedom. Following this result, then the variance of random sample of size  $n$  from a normal distribution has a  $1-\alpha\%$  CI defined by

$$\left[ \frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right]$$

$$\left[ \frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right]$$

Example 1.

In 16 test runs the fuel consumption of an experimental engine had std of 2 gallons

- 1) Construct a 99% CI for variance measuring variability of fuel consumption of the engine.

Solution

$$\frac{(16-1)2^2}{X^2} < \delta^2 < \frac{(16-1)2^2}{X^2}$$

$\xrightarrow{X = 0.8\%}$        $\xleftarrow{X = 99.5\%}$

6.32.8                          4.601

$$\frac{15 \times 4}{32.8} < \delta^2 < \frac{15 \times 4}{4.601}$$

ii) Construct 99% CI for standard deviation

Solution

$$\frac{15 \times 4}{32.8} < \delta^2 < \frac{15 \times 4}{4.601}$$

Ex:

Find the 95% CI for variance and standard deviation of nicotine content in cigarette manufactured if a sample of 20 cigarettes, and a std of 1.6 milligrams.