Inferential statistics

Population

Population is the set of all individuals, items or data of interest

Population Parameter:— A characteristic (numeric) that describes a population is referred to as a population parameter.

⇓

⇒ Whole population do not have ~~acc~~ access to ~~whole population~~ investigate

⇓

Limited to working with a sample of individuals in the population.

Inferential statistics:— characteristics of sample to infer or draw conclusions on what the unknown parameters in the given population would be.

⟶ Inferential statistics uses the sample data to draw conclusions (or inferences) about, or estimate parameters of, the environment from which the data come.

⇒ Inferential statistics is concerned with making generalizations based on a set of data by going beyond information contained in the set.

⇒ How Selected sample from population for drawing/study the population ~~charac~~ characteristics.

⇒ Leads development of sampling theory

Inferential statistics considers :-

(i) Sampling Theory :- Selecting sample from some collection that is too large to examined completely.

(ii) Estimation Theory :- Making some prediction on ~~or~~ estimate of population parameter based on the available data.

(iii) Hypothesis Testing :- Inferential procedure that uses sample data to evaluate the credibility of a hypothesis about a population.

Hypothesis Testing ≡ ~~Detection T~~ Detection Theory

≡ Decision Theory.

(iv) **Regression Analysis:**— Find mathematical expressions that best represent the collected data.

→ Sampling Theory :—

Population: Collection of data to be studied is called population.

Population can be finite or infinite.

⇒ Due to different reason it it is difficult to ~~study~~ and some time impossible to examine the entire population.

⇓

In such case study is carried out on ~~small~~ with small part of the population called **sample**

⇓

Fact about entire population can be inferred from the result obtained from the sample

⇓ Sampling

Process of obtaining samples from population is called sampling.

Sampling is ~~concern~~ related with the Selection of a subset of the members of population to estimate/infer the ~~charact~~ characteristics of the whole ~~the~~ population.

⇓

Reliability of the inference/conclusion drawn about population depends whether the sample ~~is~~ is chosen to represent the population sufficiently well.

⇓

One way to ensure sample sufficiently represent the population is each ~~sa~~ member of population has same chance of being in the sample. Samples constructed in this manner are called random samples.

Enroll
Page No: 74
Date: / /

Random Sampling :

Population Random Variable : X

Population Size : N

→ Sample size, n, is obtained ~~by~~ from population through realization of random variable X. Let $x_1, x_2, x_3, \ldots, x_n$ be the values of random variable X.

i.e. $x_1, x_2, x_3, \ldots, x_n$ is a sample

— Assume that the each value is independent of each ~~oth~~ other. This sample is realization of values as sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variable (i.i.d.), each of have same as population distribution as X.

— Random sample of size 'n' as a sequence of independent and identically distributed random variables $X_1, X_2, -- X_n$.

- once sample has been taken, the values denoted by $x_1, x_2, x_3, -- x_n$. (obtained sample)

Sample statistics :— Any quantity obtained from a sample for the purpose of estimating a population is called a sample ~~stat~~ statistics (or simply statistics).

$\hat{\theta}$ is estimator of θ is ~~function~~ of a population describe by random variable X is function of random sample $X_1, X_2, -- X_n$

$$\hat{\theta} = f(X_1, X_2, --- X_n)$$

- Most common estimators are the sample mean and the sample variance.

Sample Mean

(Ideally sample mean be mean and variance obtained from sample is same as population mean and variance.

$X_1, X_2, \ldots X_n$ is random variables for sample of size n. We define sample mean $\bar{X}$ as the following random variable

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{--- (1)}$$

X_i are random variable that are assumed to have pdf/pmf ($f_X(x)$ / $p_X(x)$) as the population random variable X.

— When sample mean, when sample values are obtained as $x_1, x_2, x_3, \ldots, x_n$, is given as

$$\boxed{\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}} \quad \text{--- (2)}$$

μ_X: mean value of population random variable X.

* Sample mean, $\bar{x}$, is random variable.
  ↳ different sample sizes gives different means

⇒ Since sample mean, $\bar{X}$, is random variable the sample mean have mean value, given by

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

$$\boxed{E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} \mu_X = \mu_X} \longrightarrow (3)$$

Since $\{X_i\}_{i=1}^{n}$ are i.i.d. and have same pdf/pmf as population random variable X.

Thus sample mean = population mean

∴ Sample mean is an unbaised estimate
⇓
Estimated value of parameter is the same as true value.

Variance of the sample mean :— Sampling is done with replacement i.e. for each draw ~~sample~~ population is not change.

Now Variance of sample mean

$$\sigma_{\bar{X}}^2 = E[(\bar{X}-\mu_X)^2] = E\left[\left(\frac{X_1+X_2+\cdots+X_n}{n} - \mu_X\right)^2\right]$$

Page No: 18
Date: / /

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} E[x_i^2] + \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} E[x_i x_j]$$
$$- \frac{2\mu_x}{n}\sum_{i=1}^{n} E[x_i] + \mu_x^2$$

Since $\{x_i\}_{i=1}^{n}$ are i.i.d. r.v. $E[x_i x_j] = E[x_i]E[x_j] = \mu_x^2$

$$E[x_i] = E[x]$$

Therefore,
$$E[x_i^2] = E[x^2]$$

$$\cancel{\sigma_{\bar{x}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} E[x_i]^2}$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2} n \cdot E[x^2] + \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} \mu_x^2$$
$$- \frac{2\mu_x}{n}\sum_{i=1}^{n}\mu_x + \mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}E[x^2] + \frac{1}{n^2}\Big[\underbrace{(n-1)\mu_x^2 + (n-1)\mu_x^2 + \dots + (n-1)\mu_x^2}_{\text{sum of } n\text{-term}}\Big]$$
$$- \frac{2\mu_x}{n}\sum_{i=1}^{n}\mu_x + \mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}E[x^2] + \frac{1}{n^2}[n(n-1)\mu_x^2] - \frac{2\mu_x}{n} n\mu_x + \mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}E[x^2] + \frac{n(n-1)\mu_x^2}{n^2} - \mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}E[x^2] + \frac{(n-1)}{n}\mu_x^2 - \mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}E[x^2] - \frac{1}{n}\mu_x^2$$

$$\sigma_{\bar{x}}^2 = \frac{1}{n}(E[x^2] - \mu_x^2)$$

$$\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}}$$

σ_x^2 : true Population Variance

4

Sampling without Replacement

Population size : N

Sample Size : n

The Variance of sample mean

$$\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}\left(\frac{N-n}{N-1}\right)}$$

Note:- Since sample mean is r.v. and obtained by sum of ~~of~~ random variables

$\Downarrow$

From central limit theorem that it tends to be asymptotically ~~a~~ normal regardless of the distribution of the random variables in samples

— In general assumption of normal distribution is true for $n \geq 30$

— Define standard normal score of the sample mean

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}}$$ then when $n \geq 30$ we

obtain

$$\boxed{F_{\bar{X}}(x) = P[\bar{X} \leq x] = \Phi\left(\frac{x - \mu_x}{\sigma_x/\sqrt{n}}\right)}$$

Ex:- Rv. X follow the exponential distribution.
With pdf $f_X(x) = 2e^{-2x}; x \geq 0$

With this distribution 36 samples are obtained.
In order to calculate sample mean $\bar{X}$.
Calculate the following

(i) $E[\bar{X}^2]$ and $E[\bar{X}^2]$
(ii) $P[\frac{1}{4} \leq \bar{X} \leq \frac{3}{4}] = ?$
i.e. Probability that sample mean lies between

$\frac{1}{4}$ and $\frac{3}{4}$.

Solution:- (i) Since X follow the exponential distribution
then true mean $E[X] = \frac{1}{2}$ and
true variance $\sigma_x^2 = \frac{1}{4}$

Since sample mean $\bar{X}$ is unbaised

$\therefore \quad E[\bar{X}] = E[X] = \frac{1}{2} = \mu_x$

(ii) $E[\bar{X}^2] = Var(\bar{X}) + \{E[\bar{X}]\}^2$
$= \sigma_{\bar{X}}^2 + \mu_{\bar{X}}^2$

$E[\bar{X}^2] = \frac{\sigma_x^2}{n} + \mu_x^2$ [Since $\mu_{\bar{X}} = \mu_X$, $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$

$= \frac{1/4}{36} + \left(\frac{1}{2}\right)^2$

$= \frac{1}{144} + \frac{1}{4}$

$E[\bar{X}^2] = \frac{37}{144}$

$Z = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}} \sim N(0,1) \rightarrow$ CDF Φ

$P[\frac{1}{4} \leq \bar{X} \leq \frac{3}{4}] = ($

$\bar{X} \sim N(\mu_x, \frac{\sigma_x^2}{n})$

$P[\frac{1}{4} < \bar{X} < \frac{3}{4}] = F_{\bar{X}}(\frac{3}{4}) - F_{\bar{X}}(\frac{1}{4})$

$= \Phi\left(\frac{\frac{3}{4} - \frac{1}{2}}{1/12}\right) - \Phi\left(\frac{\frac{1}{4} - \frac{1}{2}}{1/12}\right)$

$= \Phi(3) - \Phi(-3) = 2\Phi(3) - 1$

The Sample Variance :

Variance $\Rightarrow$ indicates the spread of values around the mean

Estimate of Variance :

The sample variance is denoted by s^2 and defined by :

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

Since $X_i - \bar{X} = (X_i - E[X]) - (\bar{X} - E[X])$

$$(X_i - \bar{X})^2 = (X_i - E[X])^2 + (\bar{X} - E[X])^2 - 2(X_i - E[X])(\bar{X} - E[X])$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - E[X])^2 + \sum_{i=1}^{n} (\bar{X} - E[X])^2 - 2(\bar{X} - E[X]) \sum_{i=1}^{n} (X_i - E[X])$$

$$= \sum_{i=1}^{n} (X_i - E[X])^2 + n(\bar{X} - E[X])^2 - 2n(\bar{X} - E[X])^2$$

$$\sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - E[X])^2 - n(\bar{X} - E[X])^2 \qquad (2)$$

Now from (1) and (2) we have

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - E[X])^2 - (\bar{X} - E[X])^2 \qquad (3)$$

Page No: 82

Now from (3) we have

$$E[s^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - E[x])^2 - (\bar{x} - E[x])^2\right]$$

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

$$= \sigma_x^2 - \frac{\sigma_x^2}{n}$$

$$\boxed{E[s^2] = \frac{(n-1)}{n}\sigma_x^2}$$

σ_x^2: Population variance

Sample variance is baised estimate

Unbaised estimate

For unbaised estimate define new r.v as

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

$$\boxed{E[\hat{s}^2] = \sigma_x^2 = \text{Population Variance}}$$

$\hat{s}^2$ is an unbaised estimate of the variance

[This is for infinite population or finite population with replacement]

Sampling With Replacement from a finite population of size N.

mean of sample variance

$$\boxed{E[s^2] = \left(\frac{N}{N-1}\right)\left(\frac{n-1}{n}\right)\sigma_x^2}$$

Sampling Distributions:

When sample size, n, is large, then from central limit theorem allows to expect sample mean follow normal distribution with standardized score given by:

$$\boxed{Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}}} \quad \text{—} (1)$$

normalization assumption is valid for $n \geq 30$

$n < 30$ ⇒ When $n < 30$, we use student's ~~t-dist~~ t-distribution by defining the normalized sample mean as:

$$T = \frac{\bar{X} - E[X]}{\hat{S}/\sqrt{n}} = \frac{\bar{X} - E[X]}{S/\sqrt{(n-1)}} \quad \text{—} (2)$$

for sample size n, the student's t-distribution is said to have n-1 degree of freedom.

The pdf of the Student's t distribution is given by:

$$f_T(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma(v/2)}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2} \quad \text{—} (3)$$

Where $\Gamma(x)$ is Gamma function of x

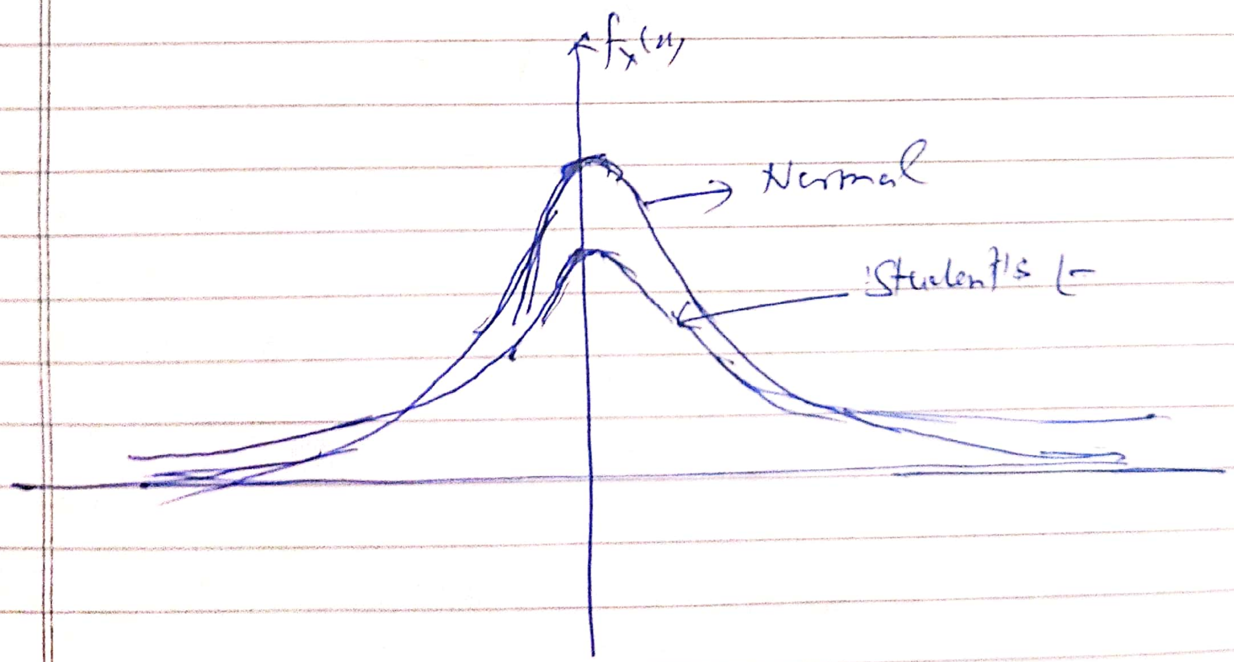
$\nu = n-1$ is the number of degrees of freedum, which is the number of independent samples.

$$\Gamma(k+1) = \begin{cases} k\Gamma k, & \text{any } k \\ k! & \text{, } k \text{ integer} \end{cases}$$

$$\Gamma(2) = \Gamma(1) = 1$$

$$\Gamma(1/2) = \sqrt{\pi}$$

Student t-distribution is heavy tailed compared to standard normal distribution



Estimation Theory

Enroll
Page No: 87
Date: / /

Estimation Theory :-

Estimation ⇒ making some prediction or estimate on population parameter based on the available sample/data.

Estimation Goal is

to estimate a variable that is not directly observable but is observed only through some other ~~variabl~~ measurable variables.

eg A random variable X has ~~distn~~ distribution that depends on some parameter θ

⇒ A statistics $g(x)$ is called estimator of θ, if any ~~value~~ observed value $x \in X$, ~~g(x)~~ $g(x)$ is considered to be an estimate of θ.

↳ Estimation of parameter can be defined as a rule or function that assigns a value to θ for each realization of x.

Page No. 21

Date / /

Another way look to Estimation

Fitting probability law / distribution of data

$\Downarrow$

where ~~data~~ the data consist of a collection of ~~set~~ realized r.v. variables generated by the probability distribution under consideration

observation data is modeled as ~~data~~ realization of i.i.d. r.v. $X_1, X_2, \ldots X_n$

with ~~$X_1, X_2, X_3 \ldots X_n \sim F(x, \theta)$~~

$X_1, X_2, X_3, \ldots, X_n \sim F(x;\theta) / F(x|\theta)$

i.e distribution depends on parameter θ that is to be estimated.

e.g. $p(x|\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, 2, \ldots$ (Poisson pmf with parameter λ.)

pdf: $f_X(x|\mu,\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$; $-\infty < x < +\infty$, Normal Distribution

Here parameter ~~$\{\mu, \sigma\}$~~ μ, σ

pdf: $f_X(x|\theta) = \theta e^{-\theta x}$; $x \geq 0$, exponential distribution with paramete θ

$$\boxed{f(x;k,\lambda) = \frac{\lambda^k x^{k-1}}{(k-1)!} e^{-\lambda x}; \quad x \geq 0}$$

Erlang distribution with parameter $\{\lambda, k\}$

~~Properties of Estimator~~ :-

⇒ The estimation of the parameter θ is the value of θ obtained from realization of x..

e.g. sample mean $\bar{x}$, is an estimator of the population ~~par~~ mean μ.

Properties of Estimator :-

① **Bias :-** Bias of estimation $\hat{\theta}$ is defined by

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$$

→ an unbaised estimator of population parameter is a statistic whose mean or expected value is equal to the parameter being estimated

Thus estimator $\hat{\theta}$ is defined ~~by~~ to be

$$\boxed{B(\hat{\theta}) = E[\hat{\theta}] - \theta}$$

unbaised if $\boxed{E[\hat{\theta}] = \theta}$ $\quad$ $\boxed{E[\hat{\theta}] = \theta}$

(ii) Efficiency :-

If sampling distributions of two statistics have the same mean, the statistic with the smaller variance is said to be more efficient estimator of the mean.

eg. let $\hat{X}_1$ and $\hat{X}_2$ are unbaised estimator of X, then $\hat{X}_1$ is more efficient estimator of X than $\hat{X}_2$ if $\sigma^2_{\hat{X}_1} < \sigma^2_{\hat{X}_2}$

(iii) Consistency :- let we ~~assume~~ use an estimator of the sample mean in the form

$$\hat{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

where x_i, $i = 1, 2, 3, \dots, n$ are observed of ~~data~~ data used to estimate θ.

" $\hat{X}$ is said to an efficient estimator of θ if it is ~~conver~~ converges in probability to θ;

Enroll

Page No: 91

Date: / /

that is

$$\lim_{n \to \infty} P[|\hat{x} - \theta| \geq \varepsilon] = 0$$

Thus as $n \to \infty$ the ~~cons~~ consistent estimator get closure to the true parameter.

Point Estimate, Interval Estimate and confidence interval:

- In Point estimate a single value is assign to the estimate
  e.g. eg sample mean, variance of sample mean

- Interval estimate :- Estimate not coincide ~~the~~ with the parameter they are estimating.
  ↳ In this parameter being estimated lies within certain interval, called "confidence ~~int~~ interval", with a certain probability.
  ↳ parameter estimate lie within two numbers, with certain degree of confidence

q - percent Confidence interval :-

↳ q% confidence interval is an interval with which the ~~prob~~ estimate ~~be~~ ~~call~~ will lie with probability $\frac{q}{100}$.

⇒ population parameter have probability $\frac{q}{100}$ belonging to the q% confidence interval.

parameter q is called the confidence level :-

eg ~~95%~~

95% confidence interval for the mean μ is a random interval that contains μ with probability 0.95.

⇒ interval ⇒ uncertainity in estimate.

Sample Mean :-

Sample mean follow the normal distribution for $n \geq 30$

For sample mean q-% confidence interval is defined as follows

Sampling with replacement from infinite population

$$\bar{X} - k\sigma_{\bar{X}} \leq \mu_X \leq \bar{X} + k\sigma_{\bar{X}}$$

population mean

Equivalently

$$\boxed{\bar{X} - \frac{k\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + \frac{k\sigma_x}{\sqrt{n}}}$$

where k is constant that depends on α and is called confidence coefficient or critical value

$k \Rightarrow$ Number of standard deviation on ~~other side~~ either side of the mean that the confidence interval is expected to cover.

Confidence limit $\bar{X} \pm k\sigma_{\bar{x}} = \bar{X} \pm \frac{k\sigma_x}{\sqrt{n}}$

and error of estimate is

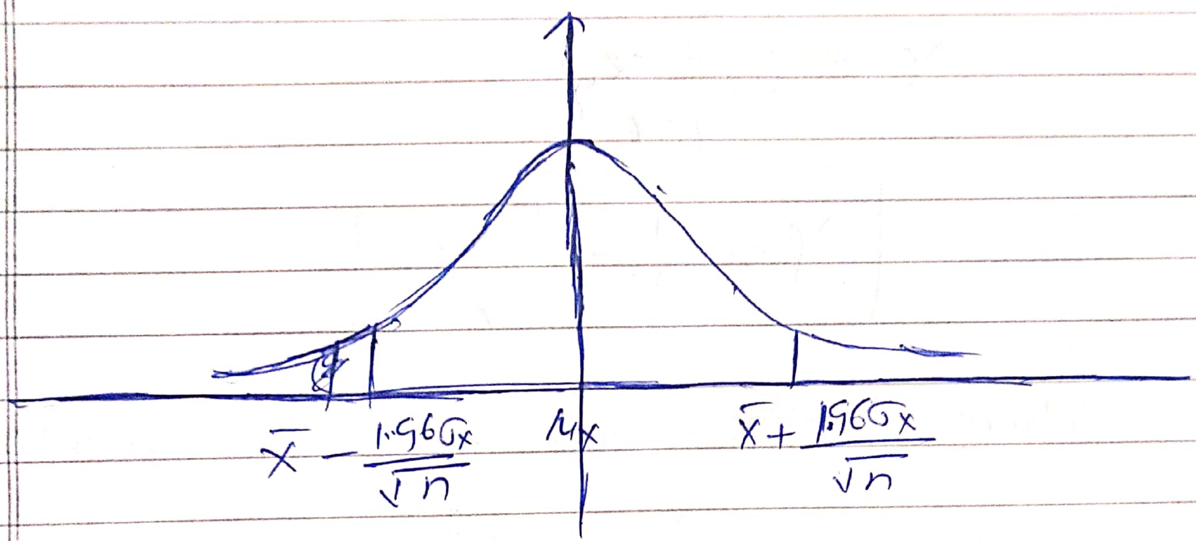
$$\boxed{k\,\sigma_{\bar{x}} = k\,\frac{\sigma_x}{\sqrt{n}}}$$

Sampling without Replacement :- Sampling without replacement from a finite population of size N; we obtain

$$\boxed{\bar{X} - \frac{k\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu_x \leq \bar{X} + \frac{k\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Table 1: Values of k for Different confidence Level

Confidence Level	99.99%	99.9%	99%	95%	90%	80%
k	3.89	3.29	2.58	1.96	1.64	1.28



→ Interval Estimate and Point Estimate :-

Interval estimate = Point estimate $\pm k_q \times$ Sample Standard deviation

k_q : Value of confidence coefficient k for specified confidence level q.

Ex: If $\sigma_x = 1$, determine the number of observations required to ensure that at the 99% confidence level, $\bar{x} - 0.1 \le E[X] \le \bar{x} + 0.1$, where

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

$$\boxed{Y = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

Solution:— The variance of the sample mean is $\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$. From table 1 for the 99% confidence level the confidence coefficient is $k = 2.58$. Thus

$$P\left[\bar{x} - \frac{2.58\sigma_x}{\sqrt{n}} \le E(X) = \mu_x \le \bar{x} + \frac{2.58\sigma_x}{\sqrt{n}}\right] = 0.99$$

$$\Rightarrow \frac{2.58\sigma_x}{\sqrt{n}} = 0.1;$$

$\sigma_x = 1$ (given)

$$\therefore \frac{2.58 \times 1}{\sqrt{n}} = 0.1$$

$$\Rightarrow \sqrt{n} = 25.8$$

$$\Rightarrow n = (25.8)^2$$

$n = (25.8)^2$

$\therefore$ $\boxed{n = 665.64}$

Thus, n must be integer, and we have

$\boxed{n = 666}$

Maximum ~~th~~ Likelihood Estimation :

To obtain best estimation of parameter(s) that characterises a random ~~variable~~ variable X

→ Choose the value(s) of parameter(s) that makes the observed values most probable

→ Let X be a r.v. whose distribution is depends on single parameter θ. Let $x_1, x_2, x_3, \ldots, x_n$ be an observed random sample.

Let X is discrete with pmf $p_X(x, \theta)$, the probability that random sample consists of these values is given by

$$L(\theta) = L(\theta;\, x_1, x_2, x_3, \ldots, x_n)$$

$$\cancel{= p_X(\theta, x_1)\, p_X(\theta_2, x_2) \ldots p_X(\theta_n}$$

$$\boxed{L(\theta) = p_X(x_1, \theta) \cdot p_X(x_2, \theta)\, p_X(x_3, \theta) \ldots p_X(x_n, \theta)}$$

Joint distribution for getting $x_1, x_2, \ldots x_n$ (1)

This ① is called likelihood function and it is function of θ, and observed values $x_1, x_2 \dots x_n$.

- If X is ~~cont~~ contn. with pdf $f_X(x)$, then the likelihood function is given by

~~$L(\theta)$~~

$$L(\theta) = f_X(x_1, \theta) \cdot f_X(x_2, \theta) \cdot f_X(x_3, \theta) \cdots f_X(x_n, \theta) \quad \text{②}$$

The maximum likelihood estimate of θ is ~~given as~~ solution of following :

$$\boxed{\hat{\theta} = \underset{\theta}{\arg\max}\ L(x_1, x_2 \dots x_n, \theta)}$$

- For k parameters : If likelihood function contains k parameters.

$$L(\theta_1, \theta_2 \dots \theta_k ; x_1, x_2, x_3 \dots x_n) = \prod_{i=1}^{n} f_X(x_i, \theta_1, \theta_2 \dots \theta_n)$$

$$\frac{\partial L}{\partial \theta_1} = 0, \quad \frac{\partial L}{\partial \theta_2} = 0 \quad \cdots \quad \frac{\partial L}{\partial \theta_k} = 0$$

Enroll
Page No: 100
Date: / /

Ex:- Suppose a random sample of size n is drawn from the Bernulli distribution. What is the maximum likelihood estimate of p, the success probability?

Soln: Let X be Bernoulli r.v. with probability of success with p. Then the pmf of X is

$$p_X(x;p) = p^x (1-p)^{1-x}, \quad x = 0, 1; \quad 0 \le p \le 1$$

The sample value $x_1, x_2, \ldots x_n$ are sequence of 0's and 1's and the likelihood function is

$$L(p; x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum x_i} (1-p)^{\sum (1-x_i)}$$

$$= p^{\sum x_i} (1-p)^{n - \sum x_i}$$

Let $y = \sum x_i$

$$\therefore L(p; x_1, x_2, x_3, \ldots, x_n) = p^y (1-p)^{n-y} \quad \text{—— (1)}$$

Page No: 101

on base e.

Now taking log^n of both side

$$\log L(p; x_1, x_2 - x_n) = y \log p + (n-y) \log(1-p)$$

Taking partial derivative we obtain:

$$\frac{\partial}{\partial p} \log L(p; x_1, x_2, x_3, \ldots, x_n) = \frac{y}{p} + \frac{(n-y)}{1-p} \times -1$$

$$= \frac{y}{p} - \frac{n-y}{1-p}$$

Now maxima of L

$$\frac{\partial}{\partial p} \log L(p; x_1, x_2, x_3, \ldots, x_n) = 0$$

$$\therefore \frac{y}{\hat{p}} - \frac{n-y}{1-\hat{p}} = 0$$

$$y(1-\hat{p}) - (n-y)\hat{p} = 0$$

$$\therefore y - \cancel{\hat{p}y} - n\hat{p} + \cancel{y\hat{p}} = 0$$

$$\therefore y - n\hat{p} = 0$$

$$\therefore \hat{p} = \frac{y}{n}$$

$$\boxed{\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

Thus $\hat{p}$ is mean of sample values.

Minimum Mean Square Error Estimation

$\hat{X}$ be an estimator of random variable X.

The estimator e is defined as

$$e = X - \hat{X}$$

e is also random variable, provides measure of how well is estimator performs.

Goodness of an estimator (How define it)?

$$\Downarrow$$

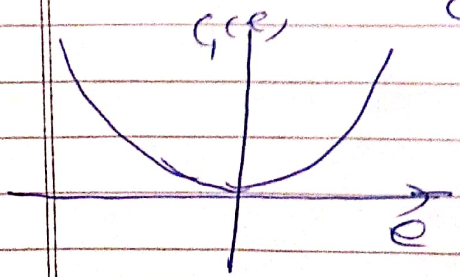
Define an appropriate cost function $C(e)$ of e.

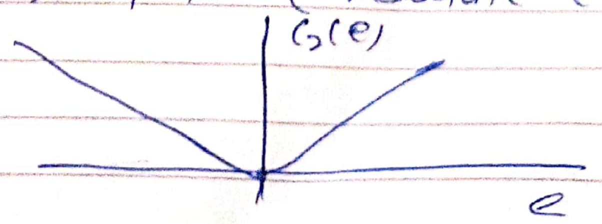
$$\Downarrow$$

Estimator is chosen such that cost $C(e)$ is minimum. The choice of cost function is subjective.

$$C_1(e) = e^2 \quad \text{(Squared error)}$$

$$C_2(e) = |e| \quad \text{(Absolute error)}$$





Mean Square Error :-

$$C(e) = E(e^2)$$

* Find the estimate in terms of first order and second order moment

Ex

Q. random variable Y estimated from the random variable X by the following linear function of X:

$$\hat{Y} = aX + b$$

Determine the value of a and b that minimize the mean square error.

Solution:-

$$e_{ms} = E[(Y-\hat{Y})^2] = E[[Y-(aX+b)]^2]$$

$$\frac{\partial e_{ms}}{\partial a} = E[2(Y-(aX+b)) \times (-X)]$$

$$= E[-2X\{Y-(aX+b)\}] = 0$$

$$\frac{\partial e_{ms}}{\partial b} = E[2(-1)\{Y-(aX+b)\}] = 0$$

$$\therefore E[XY] = aE[X^2] + bE[X]$$

$$E[Y] = aE[X] + b$$

$$\therefore a^* = \frac{E[XY] - E[X]E[Y]}{\sigma_X^2} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\rho_{XY}\sigma_X\sigma_Y}{\sigma_X^2}$$

$$\boxed{a^* = \frac{\rho_{XY}\sigma_Y}{\sigma_X}}$$

$$b^* = E[Y] - \frac{\sigma_{XY}E[X]}{\sigma_X^2}$$

$$b^* = E[Y] - \frac{\rho_{XY}\sigma_Y E[X]}{\sigma_X}$$

$$e_{mms} = e_{ms}|_{a^*} = \sigma_Y^2 - \rho_{XY}^2\sigma_Y^2 = \sigma_Y^2(1-\rho_{XY}^2)$$