# Statistical Inference

- Estimation
  - Point estimation
  - Interval estimation
- Testing of hypothesis

## Point Estimation :-

$x_1, x_2, \cdots, x_n$ is a random sample from a popn with distribution $F(x, \underline{\theta})$, $\underline{\theta} \in \Omega$.

We want to estimate a parametric $g(\underline{\theta})$ based on a statistic $T(\underline{x})$.

Then $T(\underline{x})$ is called point estimator.

where $\underline{x} = (x_1, \cdots, x_n)$.

For example, in order to estimate average height of post graduate girls students in a college, ~~The~~ a random sample of $n$ girls is taken and we use $T(\underline{x}) = \frac{1}{n}(x_1 + \cdots + x_n)$.

Or, in estimating average income of one one state we use median income $\tau(\underline{x}) = \text{med}(x_1, \cdots, x_n)$.

## ~~Cre~~ Criteria of Good Estimators

1. Unbiasedness :- A estimator $T(\underline{x})$ is said be unbiased for estimating $g(\underline{\theta})$ if

$$E_{\underline{\theta}}\{T(\underline{x})\} = g(\underline{\theta}) \quad \forall \ \underline{\theta} \in \Omega.$$

Ex:- Let $x_1, x_2, \ldots, x_n \sim P(\lambda), \lambda > 0$

Then $E(x_i) = \lambda \ \forall i = 1, 2, \ldots, n$

so, $E(\bar{x}) = \frac{1}{n} E(x_1 + \cdots + x_n) = \frac{n\lambda}{n} = \lambda$

so, $\bar{x}$ is an unbiased for estimating $\lambda$;

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

$$E(S^2) = \lambda^2$$

Te

Note:- $cS^2 + (1-c)S^2$ is also unbiased for $\lambda$.

* Let ~~E~~ $T(x) = \frac{x}{n}$.

Then $E(\frac{x}{n}) = p$

so, $\frac{x}{n}$ ~~is a~~ (i.e., sample proportion is unbiased for the popn proportion).

Theorem:- Sample mean is unbiased estimator ~~for~~ of the popn mean.

Proof:- Let $x_1, x_2, \ldots, x_n$ be a random sample from a popn with mean $\mu$ and variance $\sigma^2$.

Then $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \rightarrow$ sample mean

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i)$$

$$= \frac{1}{n} n\mu = \mu$$

$\Rightarrow$ $\bar{x}$ is an unbiased estimator of $\mu$.

Scanned by TapScanner

(ii) Sample variance is unbiased estimator of the popn variance.

Consistency :- An estimator $T_n = T(x_1, \dots x_n)$ is said to be consistent for estimating $g(\theta)$ if for every $\varepsilon > 0$

$P(|T_n - g(\theta)| < \varepsilon) \to 1$ as $n \to \infty$

or $P(|T_n - g(\theta)| > \varepsilon) \to 0$ as $n \to \infty$.

Theorem :- Sample Mean is consistent estimator of population mean, when it exists.

→ ~~$\bar{x}$ →~~ Let $\mu \to$ population mean
$\sigma^2 \to$ popn variance
& $\bar{x} \to$ sample mean.

Then $V(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^{n} var(x_i)$

$= \frac{n\sigma^2}{n^2}$ [ ~~if $x_i$~~

$= \sigma^2/n$

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

Then ~~$P(|\bar{x} - \mu| < \varepsilon)$~~ by Chebyshev inequality → $P(|\bar{x} - \mu| > \varepsilon) \le \frac{var(x)}{\varepsilon^2}$

$= \frac{\sigma^2}{n\varepsilon} \to 0$ as $n \to \infty$

$\Rightarrow P(|\bar{x} - \mu| > \varepsilon) \to 0$ as $n \to \infty$

$\Rightarrow \bar{x}$ is consistent estimator of $\mu$.

Scanned by TapScanner

## Efficiency of Estimators:-

Suppose two estimators $T_1$ and $T_2$ are given for $g(\underline{\theta})$. We define

$MSE(T_1) = E(T_1 - g(\underline{\theta}))^2$

$MSE(T_2) = E(T_2 - g(\underline{\theta}))^2$

We say that $T_1$ is more efficient than $T_2$

If $MSE(T_1) \leq MSE(T_2) \ \forall \ \underline{\theta} \in \Omega$.

If $T_1$ and $T_2$ are unbiased for $g(\underline{\theta})$, then the condn becomes $Var(T_1) \leq Var(T_2) \ \forall \underline{\theta} \in \Omega$.

## Karl Pearson :-

## Method of Moments

Let $x_1, x_2, \ldots, x_n$ be a random sample from $F(x, \underline{\theta})$, $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in \Omega$.

Find first $k$ moments of this distn -

$$\left.\begin{aligned} \mu_1' &= g_1(\theta_1, \theta_2, \ldots, \theta_k) \\ \mu_2' &= g_2(\theta_1, \theta_2, \ldots, \theta_k) \\ &\vdots \\ \mu_k' &= g_k(\theta_1, \ldots, \theta_k) \end{aligned}\right\} \cdots (1)$$

Suppose the soln of the system of equns (1) is

$$\left.\begin{aligned} \theta_1 &= h_1(\mu_1', \ldots, \mu_k') \\ &\vdots \\ \theta_k &= h_k(\mu_1', \ldots, \mu_k') \end{aligned}\right\} - (2)$$

Let $\alpha_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r \rightarrow$ rth sample moments

$r = 1, 2, \ldots, k$

Then the method of moments estimators (MMEs of $\theta_1, \theta_2, \cdots, \theta_k$ are given by

$\hat{\theta}_1 = h_1(\alpha_1, \cdots, \alpha_k), \cdots$

$\hat{\theta}_k = h_k(\alpha_1, \cdots, \alpha_k)$.

Examples:-

1. Let $X_1, X_2, \cdots, X_n \sim P(\lambda)$.

$\mu_1' = \lambda$ so, $\hat{\lambda} = \bar{X}$

So, $\bar{X}$ is the MME of $\lambda$.

2. Let $X \sim Bin(\underset{\text{known}}{n}, p)$, $0 < p < 1$.

$$\mu_1' = np \Rightarrow p = \frac{\mu_1'}{n}.$$

$$\hat{p} = \frac{X}{n}, \text{ MME of } p.$$

3. $X_1, X_2, \cdots, X_k \sim Bin(n, p)$, where both $n$ and $p$ are unknown.

$\mu_1' = np$, $\mu_2' = np(1-p) + n^2p^2$

$p = \frac{\mu_1'}{n}$ $\Rightarrow \mu_2' = \cancel{n} \cdot \frac{\mu_1'}{\cancel{n}}\left(1 - \frac{\mu_1'}{n}\right) + \frac{\cancel{n^2} \cdot \mu_1'^2}{\cancel{n^2}}$

$$\Rightarrow \frac{\mu_2' - \mu_1'^2}{\mu_1'} = 1 - \frac{\mu_1'}{n}$$

$$\Rightarrow \frac{\mu_1'}{n} = 1 - \frac{\mu_2' - \mu_1'^2}{\mu_1'} = \frac{\mu_1' - \mu_2' + \mu_1'^2}{\mu_1'}$$

$$\Rightarrow n = \frac{\mu_1'^2}{\mu_1'^2 + \mu_1' - \mu_2'}.$$

So, $p = \frac{\mu_1'^2 + \mu_1' - \mu_2'}{\mu_1'}$

So, $\hat{n} = \frac{\bar{X}^2}{\bar{X}^2 + \bar{X} - \frac{1}{n}\Sigma X_i^2} = \frac{\bar{X}^2}{\bar{X} - \frac{1}{n}\Sigma(X_i - \bar{X})^2}$

$\hat{p} = \frac{\bar{X}^2 + \bar{X} - \frac{1}{n}\Sigma X_i^2}{\bar{X}} = \frac{\bar{X} - \frac{1}{n}\Sigma(X_i - \bar{X})^2}{\bar{X}}$

Scanned by TapScanner

These ~~a~~ MME's of $n$ and $p$.

3. Let $x_1, \cdots, x_n \sim N(\mu, \sigma^2)$.

$\mu_1' = \mu, \quad \mu_2' = \mu^2 + \sigma^2$

$\sigma^2 = \mu_2' - \mu_1'^2$

$\Rightarrow \mu = \mu_1'$,

$\Rightarrow \hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \Sigma x_i^2 - \bar{x}^2$

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

These are the MMEs of $\mu$ & $\sigma^2$.

4. Let $x_1, \cdots, x_n \sim Exp(\lambda), \lambda > 0$.

$\mu_1' = \frac{1}{\lambda}$,

$\Rightarrow \lambda = \frac{1}{\mu_1'}$

$\Rightarrow \hat{\lambda} = \frac{1}{\bar{x}}$

5. ~~$x_1, \cdots, x_n \, U$~~

$x_1, x_2, \cdots, x_n \sim U(a,b), \; a < b$.

$\mu_1' = \frac{a+b}{2}, \quad \mu_2' = \frac{a^2+b^2+ab}{3}$

$a + b = 2\mu_1'$

$\Rightarrow a^2 + b^2 + 2ab = 4\mu_1'^2$

$a^2 + b^2 + ab = 3\mu_2'$

$\Rightarrow ab = 4\mu_1'^2 - 3\mu_2'$

$a^2 + b^2 - 2ab = 4\mu_1'^2 - 4(4\mu_1'^2 - 3\mu_2')$

$= 4[3\mu_2' - 3\mu_1'^2]$

$= 12(\mu_2' - \mu_1'^2)$

$\Rightarrow a - b = \pm 2\sqrt{3}\sqrt{(\mu_2' - \mu_1'^2)}$

$\Rightarrow b - a = 2\sqrt{3}\sqrt{\mu_2' - \mu_1'^2}$

$a = \mu_1' - \sqrt{3(\mu_2' - \mu_1'^2)}$

$b = \mu_1' + \sqrt{3(\mu_2' - \mu_1'^2)}$

Scanned by TapScanner

so, $\left.\begin{aligned}\hat{a} &= \bar{x} - \sqrt{\frac{3}{n}\Sigma(x_i-\bar{x})^2} \\ \hat{b} &= \bar{x} + \sqrt{\frac{3}{n}\Sigma(x_i-\bar{x})^2}\end{aligned}\right\}$ MMEs of $a$ & $b$.

Ex:- Find the MMEs for.

① $x_1, \dots, x_n \sim \text{Gamma}(n, \alpha)$

② $x_1, \dots, x_n \sim \text{Geo}(p)$

③ $x_1, \dots, x_n \sim NB(n, p)$

④ $x_1, \dots, x_n \sim \text{Beta}(m, n)$

⑤ $x_1, x_2, \dots, x_n \sim \text{Pareto}(\alpha, \beta)$.

Pareto$(\sigma, \alpha)$

$\rightarrow f(x) = \begin{cases} \frac{\alpha \sigma^\alpha}{x^{\alpha+1}}, & x>0 \\ 0. \end{cases}$

$p(x, \theta) \rightarrow P_\theta(X = x)$.

$f(x, \theta) \rightarrow$ density of $X$ at $x$.

When $\theta$ is the true parameter value.

Method of Maximum Likelihood

R.A. Fisher (1912).

Suppose an urn contains some black and white balls. It is known that the ratio of number of white balls to black balls is either 1 : 3 or 3 : 1. It is desired to estimate the proportion $p$, of white balls in the urn. We draw three balls one by one with replacement from the urn, let $X$ be the number of white balls drawn. Then

$P_p(X=x) = \binom{3}{x} p^x (1-p)^{3-x}$

$X \sim \text{Bin}(3, p)$

& $p = \frac{1}{4}$ or $\frac{3}{4}$

$\downarrow$

$(1-p)^3$.

$3p(1-p)^2$

| $X$ | 0 | 1 | 2 | 3 | |
|---|---|---|---|---|---|
| $p = 1/4$ | $\frac{27}{4}$ | $\frac{27}{64}$ | $\frac{9}{64}$ | $\frac{1}{64}$. | |
| $p = 3/4$ | $\frac{1}{64}$ | $\frac{9}{64}$ | $\frac{27}{64}$. | $\frac{27}{64}$ | |
| : | | | | | |