Estimation theory

Unit I

Point estimation - properties of estimations: Consistency and efficiency of a statistic - Simple problem

Unit II

Unbiasedness - Properties, minimum variance unbiased estimators, Raw-Blackwell theorem, Sufficiency and completeness, Lehman-Scheffe's theorem. Cramer Rao inequality - Simple problems

Unit III

Methods of estimation - Maximum likelihood estimation method. Asymptotic properties of MLE - Simple problems.

Unit IV

Interval estimation - Confidence level and confidence coefficient, Confidence, Interval for properties, difference between proportions, single mean and difference proportions - Simple problems.

Unit - V

Construction of confident interval for variance based on chi square Student's t and F distribution - Simple problems

Scanned with CamScanner

unit - I

① point estimation:

with the helps of sample observation we find the a value which is take as a value of the parameter θ. This value is termed as point estimator.

Type of estimate:

i) point estimate or simple estimate

ii) Lower & upper limits of the interval estimate.

Scanned with CamScanner

Scanned with CamScanner

Definition :

Sample / parameter

2

Statistics

Test

(Any Function of the random sample $x_1, x_2, \ldots, x_n$ That are being observed say, $T_n(x_1, x_2, \ldots x_n)$ is called statistics)

(2) Characteristic of Estimators: (properties of estimator)

The following are some of The criteria that should be statisfied by a good estimator:

i) Unbiasedness ii) consistency iii) Efficiency and iv) sufficiency

we Should now briefly explain These terms one by one.

Scanned with CamScanner

Unbiasedness :

Definition:

An estimator $T_n = T(x_1, x_2, \ldots x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if

$E(T_n) = \gamma(\theta)$, for all $\theta \in \Theta$ we have seen in chapter 13 that in sampling of a population with mean μ and variance σ^2 $E(\bar{x}) = \mu$ and $E(S^2) \neq \sigma^2$ but $E(s^2) = \sigma^2$ Hence there is a reason to prefer

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$ to the sample variance

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

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

If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$. it is said to be negatively biased The amount of bias $b(\theta)$ being given by

$b(\theta) = E(T_n) - \gamma(\theta) \cdot \theta \in \Theta$

Consistency Definition:

An estimator $T_n = T(x_1, x_2, \ldots x_n)$ based on a random sample of size n, is said to be consistent estimator of $\gamma(\theta)$, $\theta \in \Theta$, The parameter space, if T_n converges to $\gamma(\theta)$ in probability, if $T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$, In other words, T_n is a consistent estimator of $\gamma(\theta)$ if for every $\epsilon > 0$, $\eta > 0$, There exists a positive integer $n \geq m(\varepsilon, \eta)$ Such that

$$P\{|T_n - \gamma(\theta)| < \varepsilon\} \rightarrow 1 \text{ as } n \rightarrow \infty$$

$\Rightarrow P\{|T_n - \gamma(\theta)| < \varepsilon\} > 1-\eta, \; \forall\, n \geq m \ldots$ where m is a some very Large value of n.

Remarks:

1. If $x_1, x_2, \ldots x_n$ is a random sample from population with finite mean, $E x_i = \mu < \infty$, Then the Khinchine's weak law of Large numbers (W.L.L.N) we have

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \xrightarrow{P} E(x_i) = \mu \text{ as } n \rightarrow \infty$$

Hence sample mean $(\bar{x}_n)$ is always a consistent estimator of a population mean (μ)

2) obviously consistency is a property concerning the behaviour of an estimator for indefinitely large values of the sample size n. (i.e) as $n \rightarrow \infty$ nothing is regarded of its behaviour For finite n.

Scanned with CamScanner

Efficiency - Definition

If T_1 is the most efficient estimator with variance V_1 and T_2 is any other estimator with variance T_2, then the efficiency E of T_2 is defined as:

$$E = \frac{V_1}{V_2}$$

obviously, E cannot exceed unity.

If $T, T_1, T_2, \ldots T_n$ are all estimators of $\gamma(\theta)$ and variance (T) is minimum, then the efficiency E_i of T_i, $(i = 1, 2, \cdots n)$ is defined as.

$$E_i = \frac{\text{Variance } T}{\text{Var } T_i} \; ; i = 1, 2, \cdots n$$

Sufficiency:

An estimator is said to be a sufficient for any parameter, if it contains all the information in the sample regarding the parameter.

population variable / sample

If $T = t(x_1, x_2, \cdots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \cdots x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1, x_2 \cdots x_n$ given T is independent of θ then T is sufficient estimator for θ

Rao - Black Well Theorem

Theoram:

Let x and y be random variables such that

$E(y) = \mu$ and $Var(y) = \sigma y^2 > 0$

Let $E(Y/x = x) = \phi(x)$

Then

$E[\phi(x)] = \mu$ and

$Var[\phi(x)] \leq Var(y)$

Proof:

Let $f_{xy}(x,y)$ be the joint p.d.f random variables x and y, $f_1(\cdot)$ and $f_2(\cdot)$. The marginal p.d.f's of x and y respectively and $h(y/x)$ be the conditional p.d.f of y for given

$x = x$ such that $h(y/x) = \{f(x,y)/f_1(x)\}$

$$E(Y/x = x) = \int_{-\infty}^{\infty} y \cdot h(y/x)\, dy$$

$$= \int_{-\infty}^{\infty} y \frac{f(x,y)}{f(x)}\, dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y\, f(x,y)\, dy = \phi(x) \text{ (say)}$$

$$\int_{-\infty}^{\infty} y\, f(x,y)\, dy = f(x) \cdot f_1(x)$$

we observe that the conditional distribution of y given $x = x$ does not depend on the parameter μ. Hence x is sufficient statistics for μ. also

$$E\{\phi(x)\} = E\{E(y/x)\} = E(y) = \mu$$

which establishes part (i) of the theorem. Now var(y)

$$var(y) = E[y - E(y)]^2 = E(y-\mu)^2$$

$$= E(y - \phi(x) + \phi(x) - \mu)^2$$

$$= E[y - \phi(x)]^2 + E[\phi(x) - \mu]^2$$

$$+ 2E[\{y - \phi(x)\}\{\phi(x) - \mu\}]$$

The product form gives.

$$E[\{y - \phi(x)\}\{\phi(x) - \mu\}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y - \phi(x)\}\{\phi(x) - \mu\} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y - \phi(x)\}\{\phi(x) - \mu\} f_1(x)\, h(y/x)\,dx\,dy$$

But

$$\int_{-\infty}^{\infty}\{y - \phi(x)\} h(y/x)\,dy = 0$$

$$[\because E(y/x = x) = \phi(x)]$$

$$E[\{y - \phi(x)\}\{\phi(x) - \mu\}] = 0$$

we get

$$var(y) = E\{y - \phi(x)\}^2 + var(\phi(x))$$

$$var\, y \geq var[\phi(x)]$$

$$var\{\phi(x)\} \leq var\, y \quad \therefore [E(y - \phi(x))^2 \geq 0]$$

which complete the proof of the theorem.

Scanned with CamScanner

1. In random sampling from normal population $N(\mu, \sigma^2)$ Find the Maximum Likelihood estimators.

10M (or) 5Mark *.*.*

i) μ when σ^2 is known

ii) σ^2 when μ is known and

iii) The simultaneous estimation of μ and σ^2

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

$$= \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{ -\frac{1}{2\sigma^2}(x_i-\mu)^2 \right\} \right]$$

$$= \left(\frac{-1}{\sigma\sqrt{2\pi}} \right)^n \exp\left\{ -\sum_{i=1}^{n} (x_i-\mu)^2 / 2\sigma^2 \right\}$$

$$\log L = \frac{n}{2} \log(2\pi) - \frac{n}{2} \log\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2$$

Case (i)

when σ^2 is known The Likelihood equation for estimating μ is

$$\frac{\partial}{\partial\mu} \log L = 0 \Rightarrow \frac{-1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\sum_{i=1}^{n} (x_i-\mu) = 0 \Rightarrow \hat{\mu} = 1/n \sum_{i=1}^{n} x_i - \bar{x} \rightarrow (1)$$

Hence MLE for μ is the sample mean $\bar{x}$

Case (ii)

Scanned with CamScanner

when μ is known the likelihood equation for estimating σ^2 is

$$\frac{\partial}{\partial \mu} \log L = 0 \Rightarrow \frac{-1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i-\mu)(-1) = 0$$

$$\sum_{i=1}^{n} (x_i-\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \longrightarrow ①$$

Hence MLE for μ is the sample mean $\bar{x}$

$$n - \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 = 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-\mu)^2 \longrightarrow ②$$

Case (iii)

The likelihood equation for simultaneous estimation of μ and σ^2 are

$$\frac{\partial}{\partial} \log L = 0 \text{ and } \frac{\partial}{\partial \sigma^2} \log L = 0$$

Thus giving $\mu^n = \bar{x}$

$$\hat{\sigma}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-\mu^n)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-\bar{x})^2 = \sigma^2$$

The sample variance.

Important Note:

It may be pointed out here that

Though

$$E(\mu^n) = E(\bar{x}) = \mu, E(\hat{\sigma}^2) = E(S^2) \neq \sigma^2$$

Hence the maximum likelihood estimatory (MLES) need not necessarily be

Unbiasedness :

Definition:

An estimator $T_n = T(x_1, x_2, \ldots x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if

$E(T_n) = \gamma(\theta)$, for all $\theta \in \Theta$ we have seen in chapter 13 that in sampling of a population with mean μ and variance σ^2 $E(\bar{x}) = \mu$ and $E(s^2) \neq \sigma^2$ but $E(S^2) = \sigma^2$ Hence there is a reason to prefer

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

variance

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

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

If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$. it is said to be negatively biased The amount of bias $b(\theta)$ being given by

$b(\theta) = E(T_n) - \gamma(\theta)$. $\theta \in \Theta$

17

Estimate:

The constant determind through sample observation which stands for population parameter θ or a function $\gamma(\theta)$ of θ is called an estimate.

Estimator

A known function $T = t(x_1, x_2 \cdots x_n)$ of the observable variates of a random sample $x_1, x_2, \cdots x_n$ whose values are used to obtain the estimate of a parameter θ or a function of θ is called an estimator.

Scanned with CamScanner

Sufficiency :

2m
5m An estimator is said to be a sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

population variable sample

If $T = t(x_1, x_2, \cdots x_n)$ is an estimator of a parameter θ, based on a sample $x_1 x_2 \cdots x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1 x_2 \cdots x_n$ given T is independent of θ then T is sufficient estimator for θ

Theorem 10m

Cramar-Rao inaquality:

Definition:

If t is an unbiased estimator for $\gamma(\theta)$ a function of parameter θ, then

$$\mathrm{var}(t) \geq \frac{\left\{\frac{\partial}{\partial\theta} \cdot \gamma(\theta)\right\}^2}{E\left(\frac{\partial}{\partial\theta} \log L\right)^2} = \frac{\left\{\gamma'(\theta)\right\}^2}{I(\theta)}$$

where $I(\theta)$ is the information on θ supplied by the sample.

In other words, cramar-Rao inaqualiity provided a lower bound $\{\gamma'(\theta)\}^2 / I(\theta)$, to the variance of an unbiased estimator of $\gamma(\theta)$.

proof

In providing this result, we assumed that there is only a single parameter θ which is unknown.

we also take the case of continuous random variable. The case of discrete randome variables can be dealt with similarly or replacing the multiple ingrals by appropriate multiple sums.

we further make the following assumptions which are known as the Regularity conditions for cramar-Rao inequality.

Scanned with CamScanner

1. The parameter space Θ is a non-degenerate open interval on the real line $R^1 (-\infty, \infty)$ ②

2. for almost all $x = (x_1, x_2 \ldots x_n)$, and for all $\theta \in \Theta$ $\frac{\partial}{\partial\theta} L(x,\theta)$ exists, the exceptional set, if any is independent of θ

3. The range of integration is independent of the parameter θ so that $f(x,\theta)$ is differentiable under integral sign.

4. If range is not independent of θ and f is 0 at the extremes of the range, that (ie) $f(a,\theta) = 0 = f(b,\theta)$ then

$$\frac{\partial}{\partial\theta} \int_a^b f \, dx = \int_a^b \frac{\partial f}{\partial\theta} dx = f(a,\theta) \frac{\partial a}{\partial\theta} + f(b,\theta) \frac{\partial b}{\partial\theta}$$

$$\frac{\partial}{\partial\theta} \int_a^b f \, dx = \int_a^b \frac{\partial f}{\partial\theta} \, dx$$

since $f(a,\theta) = 0 = f(b,\theta)$

4) The conditions of uniform convergence of integral are satisfied so that differentiation under the integral sign is valid.

5. $I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta} \log L(x,\theta)\right\}^2\right]$, exists and is positive for all $\theta \in \Theta$

Let x be a random variable following the p.d.f $f(x,\theta)$ and let L be the likelihood function of the random sample $(x_1, x_2 \ldots x_n)$ from this population. Then $L = L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$

since L is the joint p.d.f of $(x_1, x_2 \ldots x_n)$

$$\int L(x,\theta)\,dx = 1 \text{ ; where}$$

$$\int dx = \int\int \cdots \int dx_1 \int dx_2 \cdots dx_n$$

Differentiating w.r.to θ and using regularity conditions given above we get

$$\int \frac{\partial}{\partial\theta} L\,dx = 0 \Rightarrow \int \left(\frac{\partial}{\partial\theta} \log L\right) L\,dx = 0$$

$$\Rightarrow E\left(\frac{\partial}{\partial\theta} \log L\right) = 0$$

Let $t = t(x_1, x_2, \ldots x_n)$ be an unbiased estimator of $\gamma(\theta)$ such that

$$E(t) = \gamma(\theta) \Rightarrow \int t \cdot L\,dx = \gamma(\theta)$$

differentiating w.r to θ, we get

$$\int t \cdot \frac{\partial L}{\partial\theta}\,dx = \gamma'(\theta) \Rightarrow \int t \left(\frac{\partial}{\partial\theta} \log L\right) L\,dx = \gamma'(\theta)$$

$$\Rightarrow E\left(t \cdot \frac{\partial}{\partial\theta} \log L\right) = \gamma'(\theta)$$

$$\operatorname{cov}\left(t, \frac{\partial}{\partial\theta} \log L\right) = E\left(t \cdot \frac{\partial}{\partial\theta} \log L\right) - E(t) \cdot E\left(\frac{\partial}{\partial\theta} \log L\right)$$

$$[\operatorname{cov}(x,y) = E(xy) - E(x)E(y)]$$

$$= \gamma'(\theta)$$

Scanned with CamScanner

correlation $r =$ ...

(4)

we have $\{r(x,y)\}^2 \le 1 \Rightarrow \{\operatorname{cov}(x,y)\}^2 \le \operatorname{var}(x) \cdot \operatorname{var}(y)$

$\therefore \left\{\operatorname{cov}\left(t, \frac{\partial}{\partial\theta} \log L\right)\right\}^2 \le \operatorname{var} t \cdot \operatorname{var}\left(\frac{\partial}{\partial\theta} \log L\right)$

$\Rightarrow \{\gamma'(\theta)\}^2 \le \operatorname{var} t \left[E\left(\frac{\partial}{\partial\theta} \log L\right)^2 - \left\{E\left(\frac{\partial}{\partial\theta}\log L\right)\right\}^2\right]$

$\Rightarrow \{\gamma'(\theta)\}^2 \le \underline{\operatorname{var} t} \cdot E\left\{\left(\frac{\partial}{\partial\theta} \log L\right)^2\right\}$

$$\Rightarrow \operatorname{var}(t) \ge \frac{\{\gamma'(\theta)\}^2}{E\left\{\left(\frac{\partial}{\partial\theta} \log L\right)^2\right\}}$$

1, 2, 3, 4
page.

which is cramar Rao- inequality ✓