Inference

To draw the valid conclusion or gist or essence about the characteristics under the study

Statistical Interference ⇒ The main objective of the statistical interference is to draw the valid { conclusions / findings about the characteristics under study based on the sample information in hand

The statistical inference is classified into two parts

① parametric inference

② Non-parametric inference

Statistical Inference

- Parametric
- non-parametric

Parametric ⇒ The mathematical form of population will be known except the unknown parameters

$$X \sim f(x, \theta)$$

f ↓ parametric population

θ ↳ unknown parameter

Our aim is to estimate θ

~~Atom~~

Parametric

- Estimation
  - Point
  - Interval
- Testing

Non-Parametric ⇒ In non-parametric part, the mathematical form of population is not known

$f(x)$

Distribution free part

~~consider~~ classified only in terms of testing

Parameter and Parameter Space

Any unknown characteristic of the population under study is termed as parameter. These are the value which provide the numeric summary of the population

In other words we can say that the parameter are the statistical constant which describe the data numerically.

For eg ⇒ The average weight of the children going to nursery class is 15 kg ; mean, median variance

usually the parameter is denoted by θ, This parameter θ ~~eg⇒ mean, median, variance etc.~~ may be single valued or vector valued

~~set of~~ i.e $X \sim f(x, \theta_1)$

$X \sim f(x, \theta_1, \theta_2, \ldots \theta_k)$

set of all possible values of the unknown parameter θ is called an parameter space. It is denoted by Θ

$$\Theta = \{\text{set of all possible values of } \theta\}$$

If X is a r.v having pdf or observed from the parametric population $f(x, \theta)$, then it is denoted by

$$X \sim f(x, \theta) \; ; \; \theta \in \Theta$$

θ is unknown but fixed constant and our aim is to estimate θ based on $\underline{x}$

For eg ⇒ ① A r.v $X \sim N(\mu, \sigma^2)$ then the parameter space is given by

$$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \; \sigma^2 \in \mathbb{R}^+\}$$

② If $X \sim P(\lambda)$

$$\Theta = \{\lambda : \lambda \in \mathbb{R}^+\}$$

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{\lfloor x} \; ; & x = 0, 1, 2, \ldots \infty, \; \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

Statistic

Let $x_1, x_2 \dots x_n$ be the random sample of the size n from $f(x,\theta)$; $\theta \in \Theta$ then any function of sample values $x_1, x_2 \dots x_n$ say $t(x) = t(x_1, x_2 \dots x_n)$ is called statistic. Clearly it is also a random variable. For example

$$t_1(x) = f(x_1)$$
$$t_2(x) = f(x_1, x_2)$$
$$\vdots$$
$$t_n(x) = f(x_1, x_2 \dots x_n)$$

① Sample mean

$$\bar{x} = \frac{1}{n}\sum x_i \; ; \; \bar{x} = \frac{x_1 + x_2}{2}$$

② Sample sd ,

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

$t(x) \rightarrow$ distribution
$\rightarrow$ samplistic distribution

Estimator and Estimate

Let $x_1, x_2 \dots x_n$ be a random sample of size 'n' from any population $f(x,\theta)$, $\theta \in \Theta$ and $t(\underline{x}) = t(x_1, x_2, x_3 \dots x_n)$ be the any statistic based on sample values $x_1, x_2 \dots x_n$. If the statistic $t(\underline{x})$ is used to estimate the unknown parameter θ of the population $f(x,\theta)$ is called as estimator

Any particular value of the estimator $t(\underline{x})$ is called the estimate of the parameter. If θ is the parameter then estimate of θ is denoted by $\hat{\theta}$ i.e

$$\hat{\theta} = \text{estimate of } \theta$$
$$= \text{function of sample values}$$

The estimate $(\hat{\theta})$ of the parameter θ may be most accurate, less accurate or worst

$\hat{\theta}$ — most accurate $[\hat{\theta} - \theta < \varepsilon$
— less accurate $[\hat{\theta} - \theta < \varepsilon_1 , \varepsilon_1 > \varepsilon$
— worst $[\hat{\theta} - \theta < \varepsilon_3 , \varepsilon_3 > \varepsilon_1 \, \& \, \varepsilon$

Criteria of a good estimator

A/c to professor ~~xx~~ R. A. Fisher, an estimator is taken as good estimator if it satisfy the following criterions

1. unbiasedness
2. consistency
3. Efficiency
4. Sufficiency

unbiasedness

Let $x_1, x_2 \dots x_n$ be the r.s of size n from $f(x, \theta), \theta \in \Theta$ and $t_n(x) = t(x_1, x_2 \dots x_n)$ be any statistic or estimator based on $x_1, x_2 \dots x_n$ for estimating the unknown parameter θ. The estimator $t_n(x)$ is called as unbiased estimator for the parameter θ or any ~~pa~~ function of θ say $\tau(\theta)$ if

$$\boxed{E(t_n(x)) = \tau(\theta)} \ \forall \ \theta \in \Theta \quad \text{--- (1)}$$

from eq (1), it is clear that the estimator t_n is said to be unbiased if the average value of t_n is $\tau(\theta)$

$\boxed{\text{average value of } t_n = \text{parametric value}}$

If t_n is not unbiased estimator of $\tau(\theta)$ then the condition given in (1) will not hold i.e

$$E(t_n) \neq \theta \text{ or } \tau(\theta)$$

Then t_n is called as biased estimator. Then the amount of biased is given by

$$b(t_n) = E(t_n) - \tau(\theta)$$

If (1) $E(t_n) - \tau(\theta) > 0 \Rightarrow E(t_n) > \tau(\theta)$. positively biased → overestimation

(2) $E(t_n) - \tau(\theta) < 0 \Rightarrow E(t_n) < \tau(\theta)$ negatively ⇒ under estimation

(3) $E(t_n) - \tau(\theta) = 0 \Rightarrow E(t_n) = \tau(\theta) \Rightarrow$ unbiased

MSE

Mean square error of an estimator T is given by

$$MSE(T) = E(T-\theta)^2$$

$$= E(\underbrace{T - E(T)}_{a} + \underbrace{E(T) - \theta}_{b})^2$$

$$= E(T - E(T))^2 + E(E(T) - \theta)^2 + 2E(T - E(T))(E(T) - \theta)$$

$$= Var(T) + (Bias)^2 + 0$$

$$MSE(T) = Var(T) + (Bias)^2$$

Q) Let $x_1, x_2 \dots x_n$ be the n r.s taken from $N(\mu, \sigma^2)$ then show that $\bar{x}$ is an unbiased estimator for μ.

Solution

Given that,

$$x_1, x_2 \dots x_n \sim N(\mu, \sigma^2)$$

$$x_i \sim N(\mu, \sigma^2) \quad \forall \ i = 1, 2 \dots n$$

$$\Rightarrow \left.\begin{array}{l} E(x_i) = \mu \\ V(x_i) = \sigma^2 \end{array}\right\} \forall \ i = 1, 2 \dots n \quad \text{———} \ (*)$$

we want to show that $\bar{x}$ is an unbiased for μ

Let

$$t_n = \bar{x}$$

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

Taking expectation both side, we get

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

$$= \frac{1}{n} E(x_1 + x_2 + \dots x_n) \quad \{ E(cx) = cE(x)$$

$$= \frac{1}{n} (E(x_1) + E(x_2) + \dots E(x_n))$$

by addition theorem of expectation

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

$$= \frac{1}{n} \sum_{i=1}^{n} \mu \quad (\text{by using } *)$$

$$E(t_n) = \frac{1}{n} = \frac{1}{n} \cdot n\mu$$

$$E(t_n) = \mu$$

t_n is an unbiased estimator of μ

<u>Q-1</u> Let $x_1, x_2, \ldots x_n$ be the random sample of size n from the Population $N(\mu, 1)$ then show that $t = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ be an unbiased estimator of μ^2+1

<u>Given</u> $x_1, x_2, x_3 \ldots x_n \sim N(\mu, 1)$

$$x_i \overset{iid}{\sim} N(\mu, 1) \quad \forall\ i = 1, 2, \ldots n$$

iid $\Rightarrow$ independently identically dependent

$$\left.\begin{array}{l} E(x_i) = \mu \\ V(x_i) = 1 \end{array}\right\} \forall\ i = 1, 2 \ldots n \quad \text{——— (1)}$$

we want to show that $t = \frac{1}{n}\sum x_i^2$ is an unbiased estimator for μ^2+1. Thus

$$E(t) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i)^2 \quad \text{——— (2)}$$

we know that,

$$V(x_i) = E(x_i^2) - (E(x_i))^2$$

$$1 = E(x_i^2) - \mu^2$$

$$E(x_i^2) = 1 + \mu^2 \quad \text{——— (3)}$$

From (2) and (3) we conclude that

$$E(t) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (\mu^2+1)$$

$$= \frac{n(\mu^2+1)}{n}$$

$$E(t) = \mu^2 + 1$$

Thus, t is an unbiased estimator for μ^2+1

Q.2 Let $x_1, x_2 \ldots x_5$ be R.S ~~from~~ of size 5 from $P(\lambda)$ then show that

a) $t_1 = \frac{2x_1 - x_2 + x_3}{2}$

(b) $t_2 = 4x_1 - x_2 - 2x_3$

c) $t_3 = \frac{5x_1 + 3x_2 - 2x_3 + x_4}{7}$.

are an unbiased estimator of λ and find variance

$\Rightarrow$ $x_1, x_2 \ldots x_5 \sim P(\lambda)$

$$\left.\begin{aligned} E(x_i) &= \lambda \\ V(x_i) &= \lambda \end{aligned}\right\} \forall \; i = 1, 2 \ldots 5 \quad \text{——} (1)$$

Now

$$E(t_3) = E\left(\frac{5x_1 + 3x_2 - 2x_3 + x_4}{7}\right)$$

$$= \frac{5E(x_1) + 3E(x_2) - 2E(x_3) + E(x_4)}{7}$$

$$= \frac{5\lambda + 3\lambda - 2\lambda + \lambda}{7} = \frac{7\lambda}{7} \quad (\text{from } (1))$$

$$\boxed{E(t_3) = \lambda}$$

$$E(t_2) = E(4x_1 - x_2 - 2x_3)$$

$$= 4E(x_1) - E(x_2) - 2E(x_3)$$

$$= 4\lambda - \lambda - 2\lambda$$

$$\boxed{E(t_2) = \lambda}$$

$$E(t_1) = E\left(\frac{2x_1 - x_2 + x_3}{2}\right)$$

$$= \frac{2E(x_1) - E(x_2) + E(x_3)}{2}$$

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

$$\boxed{E(t_1) = \lambda}$$

$$V(t_1) = V\left(\frac{2x_1 - x_2 + x_3}{2}\right)$$

$$= \frac{4V(x_1) + V(x_2) + V(x_3)}{4}$$

$$= \frac{4\lambda + \lambda + \lambda}{4} = \frac{6\lambda}{4} = \frac{3}{2}\lambda$$

$$V(t_2) = V(4x_1 - x_2 - 2x_3)$$

$$= 16V(x_1) + V(x_2) + 4(V(x_3))$$

$$= 16\lambda + \lambda + 4\lambda$$

$$= 21\lambda$$

$$V(t_3) = V\left(\frac{2x_1 - x_2 + x_3}{2}\right)$$

$$= \frac{1}{4}\left[4V(x_1) + V(x_2) + V(x_3)\right]$$

$$= \frac{1}{4}\left[4\lambda + \lambda + \lambda\right]$$

$$= \frac{6\lambda}{4} = \frac{3}{2}\lambda$$

Q-3 Let us consider an example of pods of peas of size $N = 5$ where the no. of peas in each pods are $x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 8$ & $x_5 = 10$. We want to draw a r.s of size $n = 2$ under the replacement and without replacement

$\boxed{x_1 \; x_2 \; x_3 \; x_4 \; x_5}$ $N = 5$ $\quad$ $\boxed{x_* \; x_*}$ $n = 2$

with r. & without r.

with replacement . $U = \{2, 4, 6, 8, 10\}$

$N = 5$

$n = 2$

under the replacement, total no of samples will be

$N^n = 5^2 = 25$

The possible r.s. are-

(2,2) (2,4) (2,6) (2,8) (2,10)

(4,2) (4,4) (4,6) (4,8) (4,10)

(6,2) (6,4) (6,6) (6,8) (6,10)

(8,2) (8,4) (8,6) (8,8) (8,10)

(10,2) (10,4) (10,6) (10,8) (10,10)

The respective mean of above sample

2	3	4	5	6
3	4	5	6	7
4	5	6	7	8
5	6	7	8	9
6	7	8	9	10

The average no of peas in pods - $\bar{x}$

$\bar{x}$	f	$f.\bar{x}$	$(\bar{x}_i - \mu_{\bar{x}})^2$	
2	1	2	16	$\mu_{\bar{x}} = \frac{1}{N}\sum f_i x_i$
3	2	6	9	$= \frac{1}{25} \times 150$
4	3	12	4	$= 6$
5	4	20	1	
6	5	30	0	
7	4	28	1	
8	3	24	4	
9	2	18	9	
10	1	10	16	
	N=25	150		

$$\mu = \bar{x} = \frac{2+4+6+8+10}{5} = 6$$

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

$$= \frac{1}{5}\left\{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2\right\}$$

$$= \frac{1}{5}\{16+4+0+4+16\} = \frac{40}{5}$$

$$\sigma^2 = 8$$

from the possible sample drawn under "with replacement" we get,

$\mu_{\bar{x}} = 6 = \mu$

= sample mean is same as the population mean

$\sigma_{\bar{x}}^2 = \frac{1}{N} \sum_i f_i (\bar{x} - \mu_{\bar{x}})^2$

$= \frac{1}{25} [1 \times 16 + 2 \times 9 + 4 \times 3 + 4 \times 1 + 5 \times 0 + 4 \times 1 + 3 \times 4 + 9 \times 2 + 16 \times 1]$

$= \frac{1}{25} [16 + 18 + 12 + 4 + 4 + 12 + 18 + 16]$

$= \frac{1}{25} [32 + 36 + 24 + 8] = \frac{1}{25} \times 100 = 4$

$\sigma_{\bar{x}}^2 = 4 < \sigma^2$

The associated variability of sample is less than the variability associated with population

<u>without replacement</u>

$U = \{2, 4, 6, 8, 10\}$

Total no of samples will be ${}^N C_n = {}^5 C_2 = 10$

The possible r.s are

(2, 4) (2, 6) (2, 8) (2, 10) (4, 6) (4, 8) (4, 10), (6, 8), (6, 10)

(8, 10)

Now the sample means are

3 4 5 6 5 6 7 7 8

9

$\mu_{\bar{x}}^{wor.} = \frac{\sum x_i}{N} = \frac{60}{10} = 6 = \mu$

$$\sigma_{\bar{x}}^{2\ wor.} = \frac{1}{10}\sum(\bar{x}_i - \mu_{\bar{x}})^2$$

$$= \frac{1}{10}\left\{(3-6)^2 + (4-6)^2 + (5-6)^2 + \dots + (9-6)^2\right\}$$

$$= \frac{1}{10}\left\{9+4+1+0+1+0+1+1+4+9\right\}$$

$$= \frac{1}{10} \times 30 = 3 < \sigma^2$$

$\sigma_{\bar{x}}^2 = 3 < \sigma^2$

$\underset{\text{(without r.)}}{\sigma_{\bar{x}}^{2\ wor}} < \sigma_{\bar{x}}^2$ (with replacement)

Hence, we see that the variability associated with "without replacement" is less than variability associated with "with replacement

__Question__

let $x_1, x_2 \dots x_n$ be the n r.s. from poisson distribution with parameter λ. Then find the __unbiased estimator__ for λ and λ^2

__Given that__

$x_1, x_2 \dots x_n \sim P(\lambda)$

$\Rightarrow x_1 \sim P(\lambda)$

$x_2 \sim P(\lambda)$

$\vdots$

$x_n \sim P(\lambda)$

$\Rightarrow x_i \overset{iid}{\sim} P(\lambda) \quad \forall\ i = 1, 2 \dots n$

$t = x_i$

$t = x_1$

$t = x_2$

$E(x_i) = \lambda$ —— (1)

$V(x_i) = \lambda$ —— (2)

from eq (1) it is clear that each x_i is an unbiased estimator for λ i.e

$E(x_1) = \lambda$

$E(x_2) = \lambda$

$\vdots$

$E(x_n) = \lambda$

Now, ~~let~~ ~~us consider~~ a statistic

Let $t = \frac{x_1 + x_2}{2}$ be any statistics based on x_1 and x_2

$$E(t) = E\left(\frac{x_1 + x_2}{2}\right)$$

$$= \frac{1}{2}\left(E(x_1) + E(x_2)\right)$$

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

$$E(t) = \lambda$$

If variance is reducing data is going towards homogeneity

$$V(t) = V\left[\frac{1}{2}(x_1 + x_2)\right]$$

$$= \frac{1}{4}\left[V(x_1) + V(x_2)\right] \quad (\because x_i\text{'s are iid})$$

$$= \frac{2\lambda}{4} = \lambda/2$$

[as it is unbiased estimator of N distribution, the variability decreases as N increases

Now again consider the mean type statistic based n r.s i.e

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

Now taking expectation both sides

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

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i) \quad \text{(by using the property of expectation)}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \lambda$$

$$E(t_n) = \frac{n\lambda}{n} = \lambda$$

$\Rightarrow t_n$ is an unbiased estimator for λ

ii) If $t = t(x_1, x_2 \dots x_n)$ be the statistic / estimator for λ^2 then

$$\boxed{E(t) = \lambda^2}$$

Let the form of above statistic t is ~~$\frac{1}{n}\Sigma$~~

$$t = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Now, for unbiasedness,

$$E(t) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{n} E(x_i^2) \quad \text{——— (1)}$$

we know that

$$V(x_i) = E(x_i^2) - (E(x_i))^2$$

$$\lambda = E(x_i^2) - \lambda^2$$

$$E(x_i^2) = \lambda + \lambda^2 \quad \text{——— (2)}$$

Now using (2) in (1)

$$E(t) = \frac{1}{n}\sum_{i=1}^{n}(\lambda^2 + \lambda)$$

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

$$\Rightarrow \quad E(t) - \lambda = \lambda^2$$

$$E(t - \lambda) = \lambda^2$$

$$E(t - \bar{x}) = \lambda^2$$

$$\Rightarrow E\left(\frac{1}{n}\Sigma x_i^2 - \frac{1}{n}\Sigma x_i\right) = \lambda^2$$

Consistency $\Rightarrow$ Large sample property

Let $x_1, x_2 \dots x_n$ be the r.s of size n from any population $f(x,\theta)$; $\theta \in \Theta$ and $T_n = t(x_1, x_2 \dots x_n)$ be any statistic/estimator based on $x_1 \dots x_n$. The estimator T_n is said to be consistent estimator for the parameter θ or function of parameter $\tau(\theta)$

If T_n converges to θ or $\tau(\theta)$ in probability for large sample i.e

$$T_n \xrightarrow{P} \theta \text{ as } n \to \infty$$

$$\Rightarrow T_n - \theta \xrightarrow{P} 0 \text{ as } n \to \infty$$

Precisely, for any $\varepsilon > 0$ & $\eta > 0$

$$P[|T_n - \theta| > \varepsilon] \to 0 \text{ as } n \to \infty$$

some large value of N

$$P[|T_n - \theta| \leq \varepsilon] \geq 1 - \eta \quad \forall\ n \geq N$$

Sufficient condition for a consistent estimator or consistency

Let $x_1, x_2 \dots x_n$ be the r.s of size n from any population - - - - - - - -

- - - - - - - - - - -

- - - - -

function of parameter $\tau(\theta)$

The estimator T_n is said to be consistent estimator for the parameter θ iff

(i) $E(T_n) \to \theta$
(ii) $V(T_n) \to \theta$
} as $n \to \theta$

$$\boxed{\lim_{n\to\infty} E(T_n) = \theta}$$

$\hookrightarrow$ Asymptotic unbiased estimator

Question

Let $x_1, x_2 \ldots x_n$ be the u r.s from the pop^n

$$f(x,\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} ; & x \geq 0 , \theta > 0 \\ 0 ; & o.w \end{cases}$$

Show that $\bar{x}$ is a consistent for θ

Given

$x_i \overset{iid}{\sim} f(x,\theta) \quad \forall i = 1,2 \ldots n$

$$f(x_i, \theta) = \begin{cases} \frac{1}{\theta} e^{-x_i/\theta} ; & x \geq 0 , \theta > 0 \\ 0 ; & o.w \end{cases}$$

for above $f(x_i, \theta)$,

$$\left.\begin{aligned} E(x_i) &= \theta \\ \text{and } v(x_i) &= \theta^2 \end{aligned}\right\} \forall i = 1,2, \ldots n$$

we want to show that

$t = \bar{x}$ is a consistent estimator of θ

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

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

$$= \frac{1}{n} \cdot n \cdot \theta = \theta$$

$$E(t) = \theta$$

$\Rightarrow t$ is an unbiased for θ

Now

$$\text{var}(t) = v\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} v(x_i)$$

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

$$\Rightarrow \lim_{n\to\infty} v(t) = \lim_{n\to\infty} \frac{\theta^2}{n} = \frac{\theta^2}{\infty} = 0$$

$$\left.\begin{aligned} E(t) &= \theta \\ v(t) &\to 0 \end{aligned}\right\} \text{as } n \to \infty$$

for large sample this implies that t is consistent estimator for the parameter θ

Let $x_1, x_2 \dots x_n$ be the r.s. of size n from $N(\mu, \sigma^2)$. Then show that $\bar{x}$ and s^2 is an unbiased ~~estim~~ and consistent estimator for μ and σ^2 $\left(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ \& } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \right)$

Solution

$$x_i \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\left.\begin{array}{l} E(x_i) = \mu \\ V(x_i) = \sigma^2 \end{array}\right\} \forall \; i = 1, 2 \dots n$$

$$\bar{x} = \frac{1}{n}\sum x_i$$

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

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

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

$\Rightarrow$ $\bar{x}$ is an unbiased estimator for μ

Now, the variance of $\bar{x}$

$$V(\bar{x}) = V\left(\frac{1}{n}\sum x_i\right)$$

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

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

For consistency

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

and $V(\bar{x}) = \frac{\sigma^2}{n}$

$$\lim_{n \to \infty} V(\bar{x}) = \lim_{n \to \infty} \frac{\sigma^2}{n} = \frac{\sigma^2}{\infty} = 0$$

$$V(\bar{x}) \to 0 \quad \text{as} \quad x \to \infty$$

Now,

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

$E(x_i - \mu)$
$(\bar{x} - \mu)$

Now, for unbiasedness

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

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

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

$$E(s^2) = \frac{1}{n-1} \sum_{i=1}^{n} E\big(\underbrace{(x_i - \mu)}_{a} - \underbrace{(\bar{x} - \mu)}_{b}\big)^2$$

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

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

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left\{ \sigma^2 - E(\bar{x} - \mu)^2 \right\}$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left\{ \sigma^2 - \frac{\sigma^2}{n} \right\}$$

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

$$\boxed{E(s^2) = \sigma^2}$$

Also, we know that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n - 1$$

$$\Rightarrow (n-1) \frac{E(s^2)}{\sigma^2} = (n-1)$$

$$E(s^2) = \sigma^2$$

If $X \sim \chi^2_{(n)}$

$E(x) = n$

$V(x) = 2n$

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma \frac{n}{2}} e^{-\chi^2/2} (\chi^2)^{n/2 - 1}, & 0 < \chi^2 < \infty \\ 0 & ; \text{ow} \end{cases}$$

$$V\left(\frac{(n-1)s^2}{\sigma^2}\right) = 2(n-1)$$

$$\frac{(n-1)^2}{\sigma^4} V(s^2) = 2(n-1)$$

$$\Rightarrow V(s^2) = \frac{2\sigma^4}{(n-1)}$$

$$\lim_{n\to\infty} V(s^2) = \lim_{n\to\infty} \left\{\frac{2\sigma^4}{(n-1)}\right\}$$

$$= \frac{2\sigma^4}{\infty}$$

$$= 0$$

s^2 is an unbiased and consistent estimator for σ^2

Question

Let $x_1, x_2 \dots x_n$ be the n r.s from the popn

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} ; & 0 < x < \theta \\ 0 ; & ow \end{cases} \qquad x_i \sim U(0, \theta)$$

Show that $\bar{x}$ is a consistent estimator for θ

Solution

$x_i \overset{iid}{\sim} U(0, \theta)$

$$\left.\begin{aligned} E(x_i) &= \frac{\theta}{2} \\ V(x_i) &= \frac{\theta^2}{12} \end{aligned}\right\} \forall\ i = 1, 2 \dots n$$

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

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

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

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\theta}{2} = \frac{1}{n} \cdot \frac{n\theta}{2} = \theta/2$$

t is unbiased for θ

Now for consistency

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

$$V(t) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \frac{\theta^2}{12}$$

$$= \frac{1}{n^2}\frac{n\theta^2}{12}$$

$$= \frac{\theta^2}{12n}$$

$$\lim_{n\to\infty} \frac{\theta^2}{12n} = \frac{\theta^2}{\infty} = 0$$

t is consistent & unbiased estimator

Invariance Property

If T_n is consistent estimator for the parameter $\tau(\theta)$. Let $\psi(\tau(\theta))$ be any continuous function of $\tau(\theta)$ then, $\psi(T_n)$ will be the consistent estimator for $\psi(\tau(\theta))$

If T_n is consistent for $\tau(\theta)$ then

$$T_n \xrightarrow{P} \tau(\theta) \text{ as } x \to \infty$$

If T_n is consistent for $\tau(\theta)$ then $\exists$ (there exist)

$$|\underbrace{\psi(T_n) - \psi(\tau(\theta))}_{A}| < \varepsilon_1 \text{ whenever } |\underbrace{T_n - \tau(\theta)}_{B}| < \varepsilon$$

$\Rightarrow B \subseteq A$

$\Rightarrow P(A) \geqslant P(B)$ (by prob law)

$$P[|\psi(T_n) - \psi(\tau(\theta))| < \varepsilon_1] \geqslant P[|T_n - \tau(\theta)| < \varepsilon] \quad \text{—— (1)}$$

We know that if T_n is consistent for $\tau(\theta)$ then for any ε and $\eta > 0$

$$P[|T_n - \tau(\theta)| < \varepsilon] > 1 - \eta \quad \forall \; n \geqslant m(\varepsilon, \eta)$$

—— (2)

using (2) in (1)

$$P[|\psi(T_n) - \psi(\tau(\theta))| < \varepsilon_1] \geqslant 1 - \eta \quad \forall \; n \geqslant m(t, \eta) \text{ as } x \to \infty$$

$\Rightarrow \psi(T_n) \xrightarrow{P} \psi(\tau(\theta))$ as $x \to \infty$

$\Rightarrow \psi(T_n)$ is a consistent for $\psi(\tau(\theta))$

Example

Example ⇒ If T_n is consistent for $\tau(\theta)$ then
$T_n \xrightarrow{P} \tau(\theta)$ as $n \to \infty$

If $x_1 \dots x_n$ be the r.s from $N(\mu, \sigma^2)$ then show that $\bar{x}$ is an consistent estimator for μ. Also construct consistent estimator for μ^2

* solve for $\bar{x}$ then by invariance property $\bar{x}^2$ is consistent estimator of μ^2

Efficiency

Consistency of an estimator depends over the sample size (n) and the variance. however, unbiasedness of an estimator does not depend over the sample size as well as variance because we may construct many unbiased estimators of same size with different variance. For example

If a r.v $X \sim N(\mu, \sigma^2)$

$x_1, x_2 \dots x_n$ of size n

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

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

$\Rightarrow V(\bar{x}) = \frac{\sigma^2}{n}$

Sample median,

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

$\Rightarrow E(\tilde{x}) = \mu$

$\Rightarrow V(\tilde{x}) = \frac{\pi}{2} \frac{\sigma^2}{n}$

⇒ $\bar{x}$ and $\tilde{x}$ both are unbiased for μ

$\Rightarrow V(\bar{x}) < V(\tilde{x})$

By above, it is clear that $\bar{x}$ is more efficient than $\tilde{x}$.

Hence, Efficiency is a property which is based on variability associated with the estimator(s)

Let T_1 and T_2 be the two estimators such that,

$$var(T_1) = V_1 \quad \& \quad var(T_2) = V_2$$

Then, efficiency (E) of an estimator T_2 wrt T_1 is defined by

$$\boxed{E = \frac{V_1}{V_2}} \qquad \boxed{E = \frac{var(T_1)}{var(T_2)}}$$

Remark

① If $E = 1$

$$\Rightarrow \frac{var(T_1)}{var(T_2)} = 1$$

$\Rightarrow$ $var(T_1) = var(T_2)$

$\Rightarrow$ Both the estimators T_1 and T_2 are equally efficient

② If $E > 1$

$$\Rightarrow \frac{var(T_1)}{var(T_2)} > 1$$

$\Rightarrow$ $var(T_1) > var(T_2)$

$\Rightarrow$ T_2 is efficient

③ If $E < 1$

$$\frac{var(T_1)}{var(T_2)} < 1$$

$\Rightarrow$ $var(T_1) < var(T_2)$

$\Rightarrow$ T_1 is more efficient

Most efficient Estimator

In class of consistent estimator, there exist an estimator which has ~~vo~~ smaller variance among all the estimators which has smaller variance among all the consistent estimators.

Let $T_1, T_2 \ldots T_n$ are the n consistent estimators. If T_3 is the most efficient estimator among $T_1, T_2 \ldots T_n$

~~Then~~, i.e $var(T_3) < var(T_i)$ $\quad i = 1, 2, 4 \ldots n$

Question

Let $x_1, x_2 \dots x_5$ be the 5 r.s from $P(\lambda)$. if

$$T_1 = x_1 + x_2 - x_3$$

$$T_2 = \frac{1}{3}(x_1 + 2x_2)$$

$$T_3 = \frac{1}{5}(x_1 + 2x_2 + x_3 - x_4)$$

$$T_4 = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$$

are the estimators based on x_i, $i = 1, 2, 3 \dots 5$

Find the efficient estimator

Solution

$x_i \overset{iid}{\sim} P(\lambda)$

$E(x_i) = \lambda = V(x_i) \quad \forall\ i = 1, 2, 3, 4, 5$

$$E(T_1) = E(x_1 + x_2 - x_3)$$
$$= E(x_1) + E(x_2) - E(x_3)$$
$$= \lambda + \lambda - \lambda$$
$$= \lambda$$

T_1 is unbiased for λ

$V(T_1) = V(x_1 + x_2 - x_3)$ all x_i are iid)

$$= V(x_1) + V(x_2) + V(x_3)$$
$$= \lambda + \lambda + \lambda$$
$$= 3\lambda$$

$$E(T_2) = \frac{1}{3}\left(E(x_1) + 2(x_2)\right)$$
$$= \frac{1}{3} \cdot 3\lambda = \lambda$$

T_2 is unbiased

$$V(T_2) = \frac{1}{9}\left[V(x_1) + 4V(x_2)\right]$$
$$= \frac{1}{9} \cdot 5\lambda = 0.55\lambda$$

$$E(T_3) = \frac{1}{5}(\lambda + 2\lambda + \lambda - \lambda)$$
$$= \frac{3\lambda}{5}$$

$\Rightarrow E(T_3) \neq \lambda$

$\Rightarrow T_3$ is an biased estimator for λ

$$V(T_3) = \frac{1}{25}(\lambda + 4\lambda + \lambda + \lambda)$$

$$= \frac{7\lambda}{25} = 0.28\lambda$$

$$T_4 = \frac{1}{5}\sum_{i=1}^{5} x_i = \bar{x}$$

$$E(T_4) = \lambda$$

$$V(T_4) = \frac{\lambda}{5} = 0.2\lambda$$

we observed that

$$V(T_4) < V(T_3) < V(T_2) < V(T_1)$$

T_4 is the most efficient estimator

$$V(T_3) = \frac{1}{25}(\lambda + 4\lambda + \lambda + \lambda)$$

$$= \frac{7\lambda}{25} = 0.28\lambda$$

$$T_4 = \frac{1}{5}\sum_{i=1}^{5} x_i = \bar{x}$$

$$E(T_4) = \lambda$$

$$v(T_4) = \frac{\lambda}{5} = 0.2\lambda$$

We observed that

$$v(T_4) < v(T_3) < v(T_2) < v(T_1)$$

T_4 is the most efficient estimator

Minimum variance unbiased Estimator (MVUE)

A statistic or estimator $T = t(x_1, x_2 \ldots x_n)$ based on r.s $x_1, x_2 \ldots x_n$ is called as MVUE for the parameter or function of parameter $\tau(\theta)$ iff

① T is an unbiased estimator for the parameter i.e

$$E(T) = \tau(\theta) \quad \forall \ \theta \in \Theta$$

② The variance of the estimator T is less than any other estimator say T_1 in class of unbiased estimator (U)

i.e $var(T) < var(T_1)$, $T \in U$

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

$$X: x_1, x_2 \ldots x_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

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

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

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

and ||ly the distribution of sample median $(\tilde{x})$

$$\tilde{x} \sim N\left(\mu, \frac{\pi}{2}\frac{\sigma^2}{n}\right)$$

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

$$v(\tilde{x}) = \frac{\pi}{2}\frac{\sigma^2}{n} \quad \Rightarrow V(\bar{x}) < v(\tilde{x})$$

we see that,

$\bar{x}$ and $\tilde{x} \in U$

and $var(\bar{x}) < var(\tilde{x})$

$\Rightarrow \bar{x}$ is a MVUE

If A & B are two events, then the conditional prob. of A given B is ~~defined~~ defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; ; \; P(B) \neq 0$$

Sufficiency

A statistic / estimator which contains all the necessary and relevant information about the parameter in the population is called as sufficient statistic / estimator

Def-2

Let $x_1, x_2 \ldots x_n$ be the random sample of size n from $f(x, \theta)$; $\theta \in \Theta$ and $T = t(x_1, x_2 \ldots x_n)$ be any statistic/ estimator based on r.s of size n. The estimator T is said to be sufficient estimator/statistic if the conditional distribution of $x_1, x_2 \ldots x_n$ or any statistic say $T_1 = t(x_1, x_2 \ldots x_n)$ given T is independent of parameter. i.e

$$f(T_1 | T) = \text{indep. of } \theta$$

Def-3 (Neyman Factorization Theorem)

Let $x_1, x_2 \ldots x_n$ be the n r.s from $f(x, \theta)$; $\theta \in \Theta$ and $T = t(x_1, x_2 \ldots x_n)$ be any statistic. The statistic T is said to be sufficient for the parameter θ. If the likelihood function of the statistic T is expressed in the following form

$$h(t(x), \theta) = g(t, \theta)\, h(\underset{\sim}{x}) \quad (\text{independent of } \theta)$$

$\hookrightarrow$ function of t and θ

Question

Let x_1, x_2 be the two r.s from $B(1,\theta)$. Then show that $x_1 + x_2$ be the sufficient statistic for θ

Solution $\quad x_1, x_2 \sim B(1,\theta)$

$$\Rightarrow x_i \sim B(1,\theta) \; ; \forall \, i = 1,2$$

$$P(X = x_i) = \begin{cases} \theta^{x_i}(1-\theta)^{1-x_i} & ; x_i = 0,1 \; ; \; 0<\theta<1 \\ 0 & ; \text{ow} \end{cases}$$

Let $T = x_1 + x_2$ and we want to prove that T is sufficient for θ

for sufficiency,

$$P[X_1 = x_1, X_2 = x_2 \mid T = x_1 + x_2] = \text{independent of } \theta$$

$$\Rightarrow T = x_1 + x_2 \sim b(2,\theta)$$

$$P[T = t] = \begin{cases} {}^2C_t \, \theta^t (1-\theta)^{2-t} & ; t = 0,1,2 \; ; \; 0<\theta<1 \\ 0 & ; \text{ow} \end{cases}$$

Now

$$P[X = x_1, X = x_2 \mid T = t] = \frac{P[X = x_1, X = x_2 \cap T = t]}{P[T = t]}$$

$$P[X = x_1, X = x_2 \mid T = x_1 + x_2] = \frac{P[X = x_1, X = x_2]}{P[T = t]}$$

$$= \frac{P[X = x_1] \cdot P[X = x_2]}{P[T = t]}$$

$$= \frac{\theta^{x_1}(1-\theta)^{1-x_1} \cdot \theta^{x_2}(1-\theta)^{1-x_2}}{{}^2C_t \, \theta^t (1-\theta)^{2-t}}$$

$$P[X = x_1, X = x_2 \mid T = t] = \frac{\theta^{x_1+x_2} \cdot (1-\theta)^{2-(x_1+x_2)}}{{}^2C_t \cdot \theta^t (1-\theta)^{2-t}}$$

$$= \frac{\cancel{\theta^t (1-\theta)^{2-t}}}{({}^2C_t) \cdot \cancel{\theta^t (1-\theta)^{2-t}}}$$

$$= \frac{1}{({}^2C_t)}$$

independent of θ

$T = x_1 + x_2$ will be sufficient for θ

<u>Question</u>

Let x_1, x_2 be the two r.s from $P(\lambda)$. Let $T_1 = x_1 + x_2$ and $T_2 = x_1 + 2x_2$ be the two statistic based on x_1 and x_2. Then show that T_1 is sufficient estimator for λ but not T_2

<u>Solution</u>

<u>Given that</u>

x_i's $\sim P(\lambda)$

$x_1 \sim P(\lambda)$

$x_2 \sim P(\lambda)$

$x_1 + x_2 \sim P(2\lambda)$

<u>for sufficiency</u>

$P[X = x_1, X = x_2 \mid T = x_1 + x_2]$ = independent of λ

$\Rightarrow T = x_1 + x_2 \sim P(2\lambda)$

$$\Rightarrow P(T=t] = \begin{cases} \frac{e^{-2\lambda}(2\lambda)^t}{\lfloor t} ; & t = 0, 1, \dots \infty \\ & \lambda > 0 \\ 0 ; & \text{o.w} \end{cases}$$

when $t = x_1 + x_2$

Now $P[X = x_1, X = x_2 \mid T = t] = \frac{P[X = x_1, X = x_2 \cap T = t]}{P[T = t]}$

$$P[X = x_1, X = x_2 \mid T = x_1 + x_2] = \frac{P[X = x_1, X = x_2]}{P[T = t]}$$

$$= \frac{P[X = x_1] \cdot P[X = x_2]}{P[T = t]}$$

$$= \frac{\frac{e^{-\lambda}\lambda^{x_1}}{\lfloor x_1} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{\lfloor x_2}}{\frac{e^{-2\lambda}(2\lambda)^t}{\lfloor t}}$$

$$P[x = x_1, x = x_2 \mid T = t] = \frac{\lfloor x_1 + x_2}{\lfloor x_1 \lfloor x_2} \cdot \frac{\cancel{e^{-2\lambda}} \cdot \cancel{\lambda^{t}}}{\cancel{e^{-2\lambda}} \cdot 2^t \cancel{\lambda^t}}$$

$$= \frac{1}{2^{(x_1+x_2)}} \cdot \frac{\lfloor x_1 + x_2}{\lfloor x_1 \lfloor x_2}$$

<u>For</u> $T_2 = x_1 + 2x_2$

x_i's $\sim P(\lambda)$

$x_1 \sim P\lambda)$

$x_2 \sim P(\lambda)$

$x_1 + 2x_2 \sim P(3\lambda)$

<u>For sufficiency</u>

$P[X = x_1, x = x_2 \mid T = x_1 + 2x_2]$ = independent of λ

$\Rightarrow T = x_1 + 2x_2 \sim P(3\lambda)$

$$\Rightarrow P[T = t] = \begin{cases} \frac{e^{-3\lambda}(3\lambda)^t}{\lfloor t} ; & t = 0, 1, 2 \ldots \infty \\ & \lambda > 0 \\ 0 ; & ow \end{cases}$$

when $t = x_1 + 2x_2$

Now

$$P[x = x_1, x = x_2 \mid T = t] = \frac{P[x = x_1, x = x_2 \cap T = t]}{P[T = t]}$$

$$= \frac{P[x = x_1, X = x_2]}{P[T = t]}$$

$$= \frac{P[X = x_1] \cdot P[X = x_2]}{P[T = t]}$$

$$\frac{\dfrac{e^{-\lambda}\lambda^{x_1}}{\lfloor x_1} \cdot \dfrac{e^{-2\lambda}\,2\lambda^{x_2}}{\lfloor 2x_2}}{e^{-}}$$

Question

Find the sufficient statistic for the parameter α in the given population

$$f(x,\alpha) = \begin{cases} \alpha . x^{\alpha-1} & ; 0<x<1 \\ 0 & ; o.w \end{cases}$$

Solution

The given parametric popⁿ is

$$f(x,\alpha) = \alpha . x^{\alpha-1} \quad , 0<x<1$$
$$\alpha > 1$$

Let $x_1, x_2 \ldots x_n$ be the n r.v.s from the above popⁿ

$x_i \sim f(x_i, \theta)$.

The likelihood function is

$$L(\underset{\sim}{x}, \theta) = f(x_1,\theta) f(x_2,\theta) \ldots f(x_n,\theta)$$
$$= \prod_{i=1}^{n} f(x_i, \theta)$$
$$= \prod_{i=1}^{n} \alpha x_i^{\alpha-1}$$
$$= \alpha^n \prod_{i=1}^{n} x_i^{\alpha-1}$$

$$\Rightarrow L(x, \theta) = \left(\alpha^n \prod_{i=1}^{n} x_i^{\alpha}\right) . \left(\prod_{i=1}^{n} x_i^{-1}\right)$$

$$= g(t_1, \alpha) . h(x)$$

$h(x)$ → independent of α → function of x_i's

$g(t_1,\alpha)$ → function of (α, x)

By Factorization thᵐ we can say that

$\prod_{i=1}^{n} x_i$ is sufficient estimator / statistic for α

Degenerate Random variable ⇒ Constant type variable

* probability concentrated at one point
* variance is zero

Complete sufficient statistic / Complete Family of Distribution

Let $x_1, x_2 \ldots x_n$ be the n r.s from the popn $f(x,\theta)$; $\theta \in \Theta$ and $t = t(x_1, x_2 \ldots x_n)$ be any statistic based on the $x_1, x_2 \ldots x_n$. No doubt the distn of t will, in general depend over the parameter θ, hence corresponding to t we have family of distribution i.e $\{g(t,\theta) ; \theta \in \Theta\}$

The statistic t or more precisely the family of distribution $\{g(t,\theta) ; \theta \in \Theta\}$ is said to be complete for any measurable function $\phi(t)$ if

$$E(\phi(t)) = 0 \Rightarrow P[\phi(t) = 0] = 1$$

$$\Rightarrow \phi(t) = 0 \text{ as everywhere}$$

$$\Rightarrow \int_t \phi(t) \cdot g(t,\theta) dt = 0$$

$$\sum_t \phi(t) g(t,\theta) = 0$$

Question ⇒ Let $x_1, x_2, \ldots x_n$ be a r.s of size n from the Population

$$f(x,\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & ; x = 0, 1 \\ 0 & ; ow \end{cases}$$

then show that the statistic $T = \sum_{i=1}^{n} x_i$ is the complete sufficient statistic for θ

Solution

Since $x_1, x_2 \ldots x_n \sim B(1,\theta)$

$$\Rightarrow x_i \overset{iid}{\sim} B(1,\theta)$$

$$\Rightarrow f(x_i,\theta) = \begin{cases} \theta^{x_i} (1-\theta)^{1-x_i} & ; x_i = 0, 1) \\ & 0 < \theta < 1 \end{cases}$$

The Likelihood function

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

$$= \theta^{\Sigma x_i} (1-\theta)^{n - \Sigma x_i} \cdot 1$$

$$= g(t,\theta) \cdot h(x)$$

By Factorization theorem

$\Rightarrow T = \sum_{i=1}^{n} x_i$ is the sufficient statistic for θ

Since $x_i \sim B(1, \theta)$

$\Rightarrow \cancel{T = \sum_{i=1}^{n} x_i}$

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

$$\Rightarrow P[T = t] = \begin{cases} {}^nC_t \theta^t (1-\theta)^{n-t} & ; t = 0, 1, \dots n \\ 0 & ; ow \end{cases}$$

Now for completeness

$E(\phi(t)) = 0$

$\Rightarrow \sum_t \phi(t)\, g(t, \theta) = 0$

$$\Rightarrow \sum_{t=0}^{n} \underbrace{\phi(t) \binom{n}{t}}_{A(t)} \theta^t (1-\theta)^{n-t} = 0$$

$$\Rightarrow \sum_{t=0}^{n} A(t) \left(\frac{\theta}{1-\theta}\right)^t (1-\theta)^n = 0$$

$$\Rightarrow A(0) \cdot (1-\theta)^n + A(1) \left(\frac{\theta}{1-\theta}\right)^1 (1-\theta)^n + \dots \dots A(n) \frac{\theta^n}{(1-\theta)^n} (1-\theta)^n = 0$$

$$\Rightarrow A(0)(1-\theta)^n + A(1) \cdot \theta (1-\theta)^{n-1} + \dots \dots A(n) \theta^n = 0$$

Now comparing the coefficients we get

$A(0) = 0 \Rightarrow \phi(1)\, {}^nC_1 = 0 \Rightarrow \phi(1) = 0$

$A(1) = 0 \Rightarrow \phi(2) = 0$

$\vdots$

$A(n) = 0 \Rightarrow \phi(n) = 0$

for any it

$\phi(t) = 0 \;\forall\; t = 0, 1, \dots n$

$\Rightarrow T = \sum_{i=1}^{n} x_i$ is complete sufficient statistic for θ

Question

Let $x_1, x_2 \dots x_n$ be a r.s of size n from $U(0, \theta)$. Then show that $T = X_{(n)} = \max(x_1, x_2 \dots x_n)$ is complete for θ

Solution

Given that

$x_1, x_2 \dots x_n \overset{iid}{\sim} U(0, \theta)$

$$f(x_i, \theta) = \begin{cases} \frac{1}{\theta} & ; 0 < x_i < \theta \\ 0 & ; ow \end{cases}$$

$F(x) = P[X \leq x]$

$= \int_0^x \frac{1}{\theta}\, dt = \frac{1}{\theta} \cdot t \Big|_0^x$

$= \frac{x}{\theta}$

Now for the distribution of $t = X_{(n)}$

$g(t) = n[F(x)]^{n-1} f(x) \Longrightarrow$ dist of max of x_i's based on ordered sample

$= n \left(\frac{x_n}{\theta}\right)^{n-1} \cdot \frac{1}{\theta}$

$= n \left(\frac{t}{\theta}\right)^{n-1} \cdot \frac{1}{\theta} \quad (\because t = x_{(n)})$

$= n \cdot \frac{t^{n-1}}{\theta^n}$

$= n \left(\frac{t}{\theta}\right)^n \cdot \left(\frac{1}{t}\right)$

$t = x_{(n)}$ is the sufficient for θ

$\Rightarrow$ For completeness

$E(\phi(t)) = 0$

$\Rightarrow \int_0^{\theta} \phi(t) \cdot g(t, \theta) = 0$

$\Rightarrow \int_0^\theta \phi(t) \cdot n t^{n-1} \theta^{-n} \cdot dt = 0$

$\Rightarrow \int_0^\theta \phi(t) \cdot t^{n-1} dt = 0$

$\Rightarrow \phi(\theta)\, \theta^{n-1} = 0$

$\Rightarrow \phi(\theta) = 0 \quad \forall \; \theta \in \Theta$

$\boxed{\phi(t) = 0}$

$t = X_{(n)}$ is complete for θ

Rao Blackwell Theorem

If an unbiased estimator and sufficient statistic for the parameter $\tau(\theta)$ exist $\exists$ a MVUE for the parameter $\tau(\theta)$ and it will be the function of sufficient statistic

Proof

Let t_1 be the unbiased estimator for the parameter $\tau(\theta)$ and t be the sufficient estimator for $\tau(\theta)$. Then by def

$E(t_1) = \tau(\theta) \quad \forall \; \theta \in \Theta$ —— (1)

Let $\phi(t) = E(t_1 | t)$ = func of sufficient statistic

Now, taking expectation both sides, we get

$E(\phi(t)) = E(E(t_1|t))$ —— (2)

We know that

$E(E(X|Y)) = E(X)$

$$\begin{aligned} E(X|Y) &= \int x \cdot f(x|y)\, dx \\ &= \int x \cdot \frac{f(x,y)}{f(y)}\, dx \\ &= \int x f(x)\, dx \end{aligned}$$

by (2)

$E(\phi(t)) = E(t_1)$

$= \tau(\theta)$ (using (1))

$E(\phi(t)) = \tau(\theta)$

$\Rightarrow \phi(t)$ is also ~~an~~ unbiased for $\tau(\theta)$

<u>Now</u> $v(t_1) = E(t_1 - E(t_1))^2$

$$= E(t_1 - \tau(\theta))^2$$

$$= E(\underbrace{t_1 - \phi(t)} + \underline{\phi(t) - \tau(\theta)})^2$$

$$= E(t_1 - \phi(t))^2 + E(\phi(t) - \tau(\theta))^2 + 2E\overset{0}{(t_1 - \phi(t))}(\phi(t) - \tau(\theta))$$

$$= E(\phi(t) - \tau(\theta))^2 + E(t_1 - \phi(t))^2$$

$$= v(\phi(t)) + \text{positive term}$$

$$v(t) > v(\phi(t))$$

$\phi(t)$ has lesser variance than t and it is unbiased for $\tau(\theta)$. Hence $\phi(t)$ is a MVUE and it is function of sufficient statistic

<u>Question</u>

Let x_1, x_2, x_3, x_4 be the r.s of size four from poisson distribution with parameter λ. If T_1, T_2 and T_3 are the three estimators based on the above samples and is defined by

$$T_1 = \frac{x_1 + x_2}{2}$$

$$T_2 = 2x_1 - x_2$$

$$T_3 = \frac{x_1 + x_2 + x_3}{3}$$

Find the minimum variance unbiased estimator for λ

<u>Solution</u>

Given that $x_i \sim P(\lambda)$ $\forall$ $i = 1, 2, 3, 4$

$$\left.\begin{aligned} E(x_i) &= \lambda \\ V(x_i) &= \lambda \end{aligned}\right\} \forall \; i = 1, 2, 3, 4$$

Now, $T_1 = \frac{1}{2}(x_1 + x_2)$

$$\Rightarrow E(T_1) = \frac{1}{2}\left(E(x_1) + E(x_2)\right)$$

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

T_1 is an unbiased for λ

$$E(T_2) = E(2x_1 - x_2)$$
$$= 2E(x_1) - E(x_2)$$
$$= 2\lambda - \lambda$$
$$= \lambda$$

T_2 is an unbiased for λ

$$E(T_3) = \frac{1}{3} E(x_1 + x_2 + x_3)$$
$$= \frac{1}{3}\left(E(x_1) + E(x_2) + E(x_3)\right)$$
$$= \frac{1}{3}(\lambda + \lambda + \lambda)$$
$$= \lambda$$

T_3 is also unbiased for λ

$T_1, T_2, T_3 \in U$ (class of unbiased estimator)

<u>Now</u>

$$V(T_1) = V\left(\frac{x_1 + x_2}{2}\right) = \frac{1}{4}\left(V(x_1) + V(x_2)\right) \quad (\because x_i\text{'s are iid})$$
$$= \frac{\lambda}{2}$$

$$V(T_2) = V(2x_1 - x_2)$$
$$= 4\lambda + \lambda = 5\lambda$$

$$V(T_3) = V\left(\frac{1}{3}(x_1 + x_2 + x_3)\right) = \frac{1}{9} \cdot 3\lambda = \frac{1}{3}\lambda$$

For any λ, we observe that $V(T_1) < V(T_2)$

$V(T_3) < V(T_2)$

$V(T_3) < V(T_1) < V(T_2)$

$\Rightarrow$ and all T_i's are unbiased

Then T_3 is a MVUE

C-R Inequality — Crammer-Rao

If t is an unbiased estimator for the parameter $\tau(\theta)$ where $\tau(\theta)$ is the function of θ, then

$$V(t) \geq \frac{(\tau'(\theta))^2}{I(\theta)}$$

$$\geq \frac{(\tau'(\theta))^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2}$$

where $I(\theta)$ is the Fisher's information based on θ supplied by sample

In order to prove the above inequality, the following regularity condition will be used

★ The parameter space Θ is non-degenerate open interval on real line $\mathbb{R}$

★ For all real x, $\frac{\partial}{\partial\theta}L(x,\theta)$ should exist

★ If the range of the integration is independent of parameters then $f(x,\theta)$ is differentiable under integral sign

If the range of the integration is dependent over the parameter then the functional value will be zero at the two extreme points i.e

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

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

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

★ The Fisher's information $I(\theta) = E\left(\frac{\partial}{\partial\theta}\log L\right)^2$ should exist for all $\theta \in \Theta$

★ If the condition of uniform convergence is satisfied then the differentiation of f is valid under integral sign

$\Rightarrow E\left(t \cdot \frac{\partial}{\partial\theta} \log L\right) = \tau'(\theta)$ — (3)

We know that

$$r^2 \leq 1$$

if x & y are the two variables then,

$$\frac{cov(x,y)^2}{V(x) \cdot V(y)} \leq 1$$

$$cov(x,y)^2 \leq V(x) \cdot V(y)$$

$$\Rightarrow \left(E(x \cdot y) - E(x) \cdot E(y)\right)^2 \leq V(x) \cdot V(y) \quad \text{— (4)}$$

If $x = t$ & $y = \frac{\partial}{\partial\theta} \log L$

by (3)

$$\left\{E\left(t \cdot \frac{\partial}{\partial\theta} \log L\right) - E(t)\, \cancel{E\left(\frac{\partial}{\partial\theta} \log L\right)}^{\,0 \text{ by } (2)}\right\}^2 \leq V(t)\, V\left(\frac{\partial}{\partial\theta} \log L\right)$$

$$\Rightarrow \left\{E\left(t \cdot \frac{\partial}{\partial\theta} \log L\right)\right\}^2 \leq V(t) \cdot V\left(\frac{\partial}{\partial\theta} \log L\right) \quad \underline{\text{using } (2)}$$

$$\Rightarrow (\tau'(\theta))^2 \leq V(t) \left\{E\left(\frac{\partial}{\partial\theta} \log L\right)^2 - \cancel{\left(E\left(\frac{\partial}{\partial\theta} \log L\right)\right)^2}^{\,0}\right\}$$

$$V(t) \cdot E\left(\frac{\partial}{\partial\theta} \log L\right)^2 \geq (\tau'(\theta))^2$$

$$\boxed{V(t) \geq \frac{(\tau'(\theta))^2}{E\left(\frac{\partial}{\partial\theta} \log L\right)^2}}$$

proved

<u>Remark</u> if $\tau(\theta) = \theta$

$$\Rightarrow \tau'(\theta) = 1$$

$$\Rightarrow \boxed{V(t) \geq \frac{1}{I(\theta)}}$$

Condition for Equality in CRI

we know that CRI is

$$v(t) \geq \frac{(\tau'(\theta))^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2}$$

The sign of equality in the above inequality will hold iff the equality sign hold for eq(I). The sign of equality will hold by cauchy schiwartz inequality

$$E(x_i, y_j)^2 \leq E(x_i)^2 \cdot E(y_j)^2$$

$$\Rightarrow \frac{x_1}{y_1} = \frac{x_2}{y_2} = \dots k$$

— (II)

If the quantity $t - \tau'(\theta)$ and $\frac{\partial}{\partial\theta}\log L$ will be proportional to each other

$$t - \tau(\theta) \propto \frac{\partial}{\partial\theta}$$

$$t - \tau(\theta) = \lambda(\theta)\frac{\partial}{\partial\theta}\log L$$

where $\lambda(\theta)$ is function of θ

$$\Rightarrow \frac{\partial}{\partial\theta}\log L = \frac{t - \tau(\theta)}{\lambda(\theta)} \quad \text{— (III)}$$

$$= A(\theta)(t - \tau(\theta)) \text{ where } A(\theta) = \frac{1}{\lambda(\theta)}$$

This is necessary and sufficient condition for an unbiased estimator to attain the lower bound to the variance of an unbiased estimator

Also in CR inequality if we take

$$\gamma^2 = 1$$

$$\Rightarrow v(t) = \frac{(\tau'(\theta))^2}{E\left(\frac{\partial}{\partial\theta}\log L\right)^2} \quad \text{— (IV)}$$

Now squaring both sides of eq^n (iii) and taking expectation

$$E\left(\frac{\partial}{\partial\theta}\log L\right)^2 = E\left[A(\theta)(t-\tau(\theta))\right]^2$$

$$= A(\theta)^2 E(t-\tau(\theta))^2$$

$$= A(\theta)\, v(t)$$

Then by eq (i) we get

$$v(t) = \frac{(\tau'(\theta))^2}{A(\theta)^2 v(t)}$$

$$(v(t))^2 = \left(\frac{\tau'(\theta)}{A(\theta}\right)^2$$

$$(v(t))^2 = [\tau'(\theta)\lambda(\theta)]^2$$

$$v(t) = |\lambda(\theta)^2 \tau'(\theta)|$$

This is the required expression for the variance of MVBUE (Minimum variance Bound unbiased Estimator)

<u>Question</u>

Let $x_1, x_2, x_3 \ldots x_n$ be the rs of size n from $N(\mu, \sigma^2)$

Find MVBE for μ, given that

x_i's $\overset{iid}{\sim} N(\mu, \sigma^2)$

$$f(x,\mu) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} & ; \quad -\infty < x_i < \infty,\ -\infty < \mu < \infty,\ \sigma > 0 \\ 0 & , \quad o.w \end{cases}$$

Now, the likelihood function is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

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

Now taking log both sides, we get

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

Now differentiating wrt μ we get

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

$$\frac{\partial}{\partial\mu}\log L = \frac{1}{\sigma^2}\Sigma(x_i-\mu)$$

$$= \frac{\Sigma x_i - n\mu}{\sigma^2}$$

$$= \frac{n\bar{x} - n\mu}{\sigma^2}$$

$$\frac{\partial}{\partial\mu}\log L = \frac{\bar{x}-\mu}{\sigma^2/n}$$

By comparing the above expression with

$$\frac{\partial}{\partial\theta}\log L = \frac{t - \tau(\theta)}{\lambda(\theta)}$$

$\Rightarrow$ $\bar{x}$ is MVBE for μ and $V(\bar{x}) = \sigma^2/n$

<u>Question</u>

A random variable x has following prob. function

$$P(X=x) = \begin{cases} \frac{e^{-\theta}\theta^x}{\lfloor x} & ; x = 0,1,2\ldots\infty \\ & \quad \theta > 0 \\ 0 & ; \text{ow} \end{cases}$$

Find MVBE for the parameter θ. Also compute the variance of MVB

<u>Solution</u>

<u>given</u> $P(X=x) = \begin{cases} \frac{e^{-\theta}\theta^x}{\lfloor x} & , x = 0,1,2\ldots n \\ 0 & , \text{ow} \end{cases}$

Now likelihood function

$$L = \prod_{i=1}^{n} P(X=x_i)$$

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

$$L = \frac{e^{-n\theta}\,\theta^{\Sigma x_i}}{\prod_{i=1}^{n} \lfloor x_i} \quad \text{——} \; (1)$$

Now taking log both sides in eq (1)

$$\log L = -n\theta \log e + \Sigma x_i \log\theta - \log\left(\prod_{i=1}^{n} \lfloor x_i\right)$$

$$\log L = \cancel{-\log\left(\prod_{i=1}^{n} \lfloor x_i\right.} - n\theta + \Sigma x_i \log\theta - \sum_{i=1}^{n}(\log x_i)$$

diff wrt parameter θ

$$\frac{\partial}{\partial\theta}\log L = \cancel{x} - n + \frac{\Sigma x_i}{\theta} - 0$$

$$\frac{\partial}{\partial\theta}\log L = \frac{\Sigma x_i - n\theta}{\theta}$$

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

$$= \frac{n(\bar{x}-\theta)}{\theta}$$

$$\frac{\partial}{\partial\theta}\log L = \frac{\bar{x}-\theta}{\theta/n}$$

Now on comparing with

$$\frac{\partial}{\partial\theta}\log L = \frac{t - \tau(\theta)}{\lambda(\theta)}$$

$\bar{x}$ is MVBE for θ and variance $= \frac{\theta}{n}$

$$V(t) = |\tau'(\theta) \cdot \lambda(\theta)|$$

$$\tau'(\theta) = \theta$$

$$\tau'(\theta) = 1$$

$$V(t) = (\lambda(\theta))$$

$$V(\bar{x}) = \theta/n$$

Question

uniform min variance unbi estimator

unique best estimator

Let $X \sim P(\lambda)$. Find UMVUE for λ and λ^2

Solution

Let λ : $x_1, x_2 \dots x_n$ be the r.s of size n from $P(\lambda)$

$$\Rightarrow x_i \sim P(\lambda) \quad \forall \quad i = 1, 2 \dots n$$

$$\Rightarrow \quad E(x_i) = V(x_i) = \lambda$$

we know that,

$$T = \sum x_i \sim P(n\lambda)$$

$$P[T = t] = \frac{e^{-n\lambda} \lambda^t}{\lfloor t} \quad t = 0, 1 \quad \lambda > 0$$

For sufficiency

$$L = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{\lfloor x_i} = \frac{e^{-n\lambda} \cdot \lambda^{\sum x_i}}{\prod \lfloor x_i}$$

$$= \lambda^{\sum x_i} \cdot e^{-n\lambda} \frac{1}{\prod \lfloor x_i}$$

$$= g(t, x) \; h(\underline{x})$$

$t = \sum x_i$ will be sufficient for λ

For completeness

$$E(\phi(t)) = 0$$

$$\Rightarrow \sum \phi(t) P[T = t] = 0$$

$$\sum_{t=0}^{\infty} \phi(t) \frac{e^{-n\lambda} \lambda^t}{\lfloor t} = 0$$

$$\Rightarrow \boxed{\phi(t) = 0}$$

$T = \sum x_i$ is complete sufficient statistic for λ

$$T \sim P(n, \lambda)$$

$$E(T) = n\lambda \quad ; \quad V(T) = n\lambda$$

$$E\left(\frac{T}{n}\right) = \lambda$$

$$\Rightarrow E\left(\frac{\Sigma x_i}{n}\right) = \lambda$$

$$\Rightarrow E(\bar{x}) = \lambda$$

$\frac{T}{n} = \bar{x}$ is UMVUE for λ

$$V(T) = E(T^2) - (E(T))^2 = n\lambda$$

$$\Rightarrow E(T^2) - (n\lambda)^2 = n\lambda$$

$$\Rightarrow E(T^2) - n^2\lambda^2 = E(T)$$

$$\Rightarrow E(T^2) - E(T) = n^2\lambda^2$$

$$\Rightarrow E(T^2 - T) = n^2\lambda^2$$

$$\Rightarrow E\left(\frac{T^2 - T}{n^2}\right) = \lambda^2$$

$\Rightarrow \frac{T(T-1)}{n^2}$ is UMVUE for λ^2

<u>Question</u>

Let $x_1, x_2 \ldots x_n$ be a r.s of size n from the population

$$P[X=x] = \begin{cases} p^x(1-p)^{1-x} & ; x = 0, 1 \\ & \quad 0 < p < 1 \\ 0 & ; \text{ow} \end{cases}$$

Find UMVUE for p and p^2

Method of Estimation

Let X be a r.v observed from the parametric family $f(x,\theta)$; $\theta \in \Theta$
i.e

$$X \sim f(x,\overset{\text{unknown but fixed}}{\theta}) \; ; \; \theta \in \Theta$$

$\hookrightarrow$ prob distribution

$$\Rightarrow X \sim f(x,\theta_1,\theta_2 \dots \theta_k) \; ; \; \theta_i \in \Theta$$

$\hookrightarrow$ case of multiparameter

the form of $f(x,\theta)$ is known except the unknown parameter θ. Hence our aim is to estimate this unknown characteristics or parameter based on information supplied by sample. It means that we are intended to find $\hat{\theta}$ i.e estimate of θ

$$\hat{\theta} = f(x_1, x_2 \dots x_n)$$

In order to get ~~the~~ $\hat{\theta}$ the following classical methods of estimation may be applied

1. Method of Maximum Likelihood Estimation (MLE)
2. Method of Moment
3. Method of minimum variance
4. Method of least square
5. Method of minimum chi-square

1. MLE Method

This is the most popular method of classical estimation of the parameter. This method is based on the likelihood function

Likelihood Function

Let $x_1, x_2 \dots x_n$ be the random sample of size n from any pop^n $f(x,\theta)$; $\theta \in \Theta$. Then the likelihood function of the observed sample $x_1, x_2 \dots x_n$ is denoted by $L(x,\theta)$ or $L(\theta)$

$$L(\theta) = \text{function of parameter}$$

$$\Rightarrow x_i \overset{iid}{\sim} f(x_i, \theta)$$

$$\Rightarrow L(\theta) = f(x_1,\theta) \cdot f(x_2,\theta) \dots \cdot f(x_n,\theta) \qquad \text{estimate of } \theta$$

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

L.F is the joint density function

Maximum Likelihood Estimator

The principal of MLE consist of or state that $\hat{\theta}$ will be maximum likelihood estimator for the parameter θ. If $\exists$ $\hat{\theta}$ i.e function of sample values which maximizes the likelihood theorem

$\Rightarrow L(\hat{\theta}) > L(\theta)$

$\Rightarrow \sup_{\theta \in \Theta} L(\theta) \geq L(\theta)$ Sup $\Rightarrow$ supremum

In order to maximize the likelihood function we use the theory of maxima and minima

$$\frac{dL(\theta)}{d\theta} = 0 \quad \text{———} \star$$

and $$\left.\frac{d^2 L(\theta)}{d\theta^2}\right|_{\theta = \hat{\theta}} < 0$$

where $\hat{\theta}$ is the value which is obtained by solving the eq $\star$

For computation of MLE

$$\left.\begin{aligned} &\frac{dL}{d\theta} = 0 \\ &\text{under the condition that} \\ &\left.\frac{d^2 L}{d\theta}\right|_{\theta=\hat{\theta}} < 0 \end{aligned}\right\} \text{———} \star$$

The above eqn implies that the function should be concave type. Also instead of using likelihood function one can use logrithm of likelihood function for simplicity or mathematical ease because log L is a monotone function. Then $\star$ may be written as

$$\left.\begin{aligned} &\frac{d}{d\theta^2} \log L = 0 \\ &\left.\frac{d^2}{d\theta^2} \log L\right|_{\theta=\hat{\theta}} < 0 \end{aligned}\right\} \text{———} (\star\star)$$

The expression given in ($\star\star$) is called as likelihood eq^n and solution of the above eq^n is called as MLE

Multparameter case

Let X be a r.v such that $X \sim f(x, \underset{\sim}{\theta})$ where the parameter $\underset{\sim}{\theta}$ is a vector quantity consisting K components/parameters i.e $X \sim f(x, \theta_1, \theta_2 \dots \theta_K)$

Now we are interested to compute MLE of $\theta_i = 1, 2 \dots k$ then we use the same approach as we discussed in the case of single parameter. Then the likelihood function is

$$L(\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(x_i, \theta_1, \theta_2 \cdots \theta_K)$$

taking log both sides, we get

$$\Rightarrow \log L(\underset{\sim}{\theta}) = \sum_{i=1}^{n} \log f(x_i, \theta_1, \theta_2 \dots \theta_K)$$

Now for MLE

$$\Rightarrow \begin{cases} \frac{d}{d\theta_1} \log L = 0 \\ \vdots \\ \frac{d}{d\theta_K} \log L \end{cases}$$

$$\Rightarrow \frac{d}{d\theta_i} \log L = 0 \quad \forall \; i = 1, 2, \dots K$$

$$\left. \frac{d^2}{d\theta_i \, d\theta_j} \log L \right|_{(\hat{\theta}_i, \hat{\theta}_j)} < 0$$

The above eq^n can be written in the form of matrix

$$\begin{bmatrix} \frac{\partial^2 \log L}{\partial \theta_1^2} & \frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_K} \\ \frac{\partial^2 \log L}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \log L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \log L}{\partial \theta_2 \partial \theta_K} \\ \frac{\partial^2 \log L}{\partial \theta_K \partial \theta_1} & \frac{\partial^2 \log L}{\partial \theta_K \partial \theta_2} & \cdots & \frac{\partial^2 \log L}{\partial \theta_K^2} \end{bmatrix}_{K \times K}$$

~~For the existence of MLE~~

The above matrix of order $k \times k$ is called as Hessian matrix which is constituted with the help of double derivatives of logrithm of likelihood function

For the existence of MLEs, the above matrix should be negative definite

Question

~~Let~~ A random variable X has the following probability distribution

$$f(x, \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}} & ; x, \mu \in R \\ & \sigma \in R^+ \\ 0 & ; \text{ow} \end{cases}$$

compute MLE of the parameter

a) μ when σ^2 is known

b) σ^2 when μ is known

c) μ and σ^2 both are unknown

Solution

The given prob f^n is

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad ; \quad x, \mu \in \mathbb{R}, \ \sigma \in \mathbb{R}^+$$

Now the likelihood function based on n r.s $x_1, x_2 \dots x_n$

$$L = L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x, \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

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

Now taking log both sides we get

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

Now for MLE of μ

$$\frac{\partial}{\partial \mu}\log L = 0 \quad \& \quad \frac{\partial^2}{\partial \mu^2}\log L\Big|_{\mu=\hat{\mu}} < 0$$

$\Rightarrow$ Now by eq①

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

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

$$\Rightarrow \quad \sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i - n\mu = 0$$

$$\Rightarrow \mu = \frac{\Sigma x_i}{n}$$

$$= \bar{x}$$

ML estimate of μ is $\bar{x}$ i.e.

$$\boxed{\hat{\mu} = \bar{x}}$$

Now, $\frac{d^2}{d\mu^2}\log L\Big|_{\mu=\hat{\mu}} = -\frac{n}{\sigma^2} < 0$

Thus, we can say that $\hat{\mu} = \bar{x}$ is MLE of μ

Now for σ^2 by eq①

$$\frac{d}{d\sigma^2}\log L = 0 - \frac{n}{2\sigma^2} + \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^4}$$

$$\Rightarrow \frac{d^2}{d\sigma^2 d\sigma^2}\log L = \frac{n}{2\sigma^4} - \frac{1}{\cancel{2}}\cdot \cancel{2}\,\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^6}$$

$$= \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^6}$$

Now for MLE of σ^2

$$\frac{d}{d\sigma^2}\log L = 0 \ \& \ \frac{\partial^2}{\partial\sigma^2\partial\sigma^2}\bigg|_{\sigma^2=\hat{\sigma}^2} < 0$$

$$\Rightarrow \frac{n}{\cancel{2}\sigma^2} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\cancel{2}\sigma^4}$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\Sigma(x_i-\mu)^2 \Longrightarrow \hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-\mu)^2$$

when μ is known

$$\hat{\mu} = 2,$$

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma(x_i-2)^2$$

MLE of σ^2 is

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

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

check

$$\frac{d^2\log L}{d\sigma^4}\bigg|_{\hat{\sigma}^2} < 0 \ ?$$

Question

Let a r.v $X \sim P(\lambda)$. Find MLE of the parameter λ and λ^2

Solution

Given that $X \sim P(\lambda)$

$$\Rightarrow P[X=x] = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & ; \ x = 0,1,\dots\infty, \ \lambda > 0 \\ 0 & ; \ \text{ow} \end{cases}$$

The likelihood function is

$$L = \prod_{i=1}^{n} P[X=x_i] = \prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{x_i}}{\lfloor x_i}$$

$$= \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n}\lfloor x_i}$$

taking log both sides

$\log L = -n\lambda + \sum_{i=1}^{n} x_i \log\lambda - \sum_{i=1}^{n} \log(\lfloor x_i)$

diff the above function wrt λ

$\frac{d}{d\lambda} \log L = -n + \sum_{i=1}^{n} x_i \frac{1}{\lambda} - 0$

diff the above f^n wrt λ again

$\frac{d^2}{d\lambda^2} \log L = 0 + \left(-\frac{\sum x_i}{\lambda^2}\right) = -\frac{\sum x_i}{\lambda^2}$

for MLE

$\frac{d}{d\lambda} \log L = 0$

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

$\Rightarrow \frac{\sum x_i}{\lambda} = n$

$\Rightarrow \lambda = \frac{\sum x_i}{n}$

$\boxed{\lambda = \bar{x}}$

$\boxed{\therefore \hat{\lambda} = \bar{x}}$ where $\hat{\lambda}$ is MLE of λ

$\frac{d^2}{d\lambda^2} \log L \Big|_{\lambda=\hat{\lambda}} = -\frac{\sum x_i}{\lambda^2}$

$= -\frac{n\bar{x}}{\hat{\lambda}^2}$

$\Rightarrow -\frac{n\bar{x}}{\bar{x}^2}$

$\Rightarrow -\frac{n}{\bar{x}} < 0$

$\therefore \bar{x}$ is MLE of λ

Since $\bar{x}$ is the MLE of λ and $\psi(\lambda) = \lambda^2$ is any function of λ then by invariance property of MLE $\bar{x}^2$ will be MLE of λ^2

<u>Question</u>

Let a r.v $X \sim P(\lambda)$. Find MLE of the parameter λ & λ^2

<u>Solution</u>

Given that $X \sim P(\lambda)$ | $\lambda^2 = \theta$, $\lambda = \sqrt{\theta}$

$$\Rightarrow P[X = x] = \begin{cases} \dfrac{e^{-\sqrt{\theta}} (\sqrt{\theta})^x}{\lfloor x} & ; x = 0, 1, \cdots \infty \\ 0 & ; ow \end{cases}$$

The likelihood function is

$$L = \prod_{i=1}^{n} P[X = x_i] = \prod_{i=1}^{n} \frac{e^{-\sqrt{\theta}} \cdot \sqrt{\theta}^{x_i}}{\lfloor x_i}$$

$$= \frac{e^{-n\sqrt{\theta}} \cdot \sqrt{\theta}^{\Sigma x_i}}{\prod_{i=1}^{n} \lfloor x_i}.$$

taking log both sides we get

$$\log L = -n\sqrt{\theta} + \sum_{i=1}^{n} x_i \log\sqrt{\theta} - \sum_{i=1}^{n} \log(\lfloor x_i)$$

diff wrt θ

$$\frac{d}{d\theta} \log L = \frac{-n\,\theta^{-1/2}}{2} + \frac{1}{2} \frac{\Sigma x_i}{\sqrt{\theta}} \cdot \theta^{-1/2} = 0$$

$$\Rightarrow \frac{n}{\sqrt{\theta}} = \frac{n\bar{x}}{\sqrt{\theta}} \cdot \frac{1}{\sqrt{\theta}}$$

$$\Rightarrow \boxed{\theta = \bar{x}^2}$$

$$\hat{\theta} = \bar{x}^2$$

Question

Let $x \sim U(0,\theta)$. Find the MLE of θ given that

$\Rightarrow$ $x \sim U(0,\theta)$

$$f(x,\theta) = \begin{cases} \frac{1}{\theta} & ; x \in (0,\theta) \\ 0 & ; \text{ow} \end{cases}$$

Likelihood function is

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

taking log both sides

$\log L = -n \log \theta$

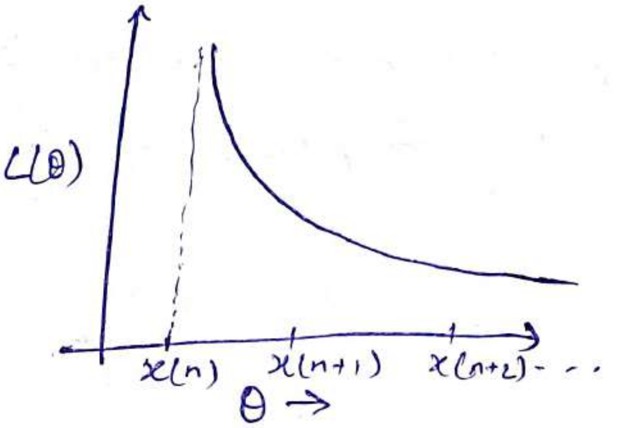
diff wrt θ

$\frac{d}{d\theta} \log L = -\frac{n}{\theta}$

if $\frac{d}{d\theta} \log L = 0$

$-\frac{n}{\theta} = 0$

$\hat{\theta} \neq$ (undefined)



$\therefore$ Range $= 0 < x < \theta$

we will focus on $0 < x_{(1)} < x_{(2)} < \dots\dots x_{(n)} < \theta$

$\Rightarrow \theta > x_{(n)}$

$L(\theta)$ will be maximized when $\hat{\theta} = x_{(n)}$

Question

If $x \sim U(\theta - 1/2 , \theta + 1/2)$ MLE of θ ?

Solution

$$x \sim U\left(\theta - \frac{1}{2} , \theta + \frac{1}{2}\right)$$

$$\Rightarrow f(x, \theta) = \begin{cases} 1 & ; \ \theta - 1/2 < x < \theta + 1/2 \\ 0 & ; \ ow \end{cases}$$

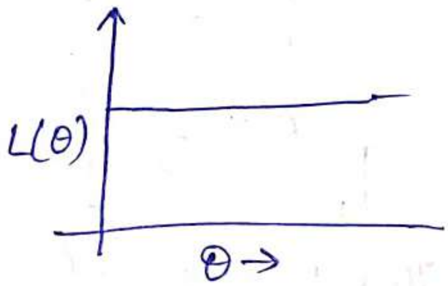
$\Rightarrow$ based on n ordered sample

$$\theta - 1/2 < x_{(1)} < x_{(2)} \cdots x_{(n)} < \theta + 1/2$$

$$x_{(1)} > \theta - 1/2 \quad \Big| \quad x_{(n)} < \theta + 1/2$$

$$x_{(1)} + \frac{1}{2} > \theta \quad \Big| \quad x_{(n)} - 1/2 < \theta$$

$$x_{(n)} - \frac{1}{2} < \theta < x_{(1)} + \frac{1}{2}$$



$$L(\theta) = 1 = \text{constant}$$

Hence for each $x_{(i)}$ it is maximum

Question

Let $x_1 = 0.5$, $x_2 = 01.3$, $x_3 = 0.85$, $x_4 = 2.1$ & $x_5 = 0.4$ be a r.s of size 5 from the popⁿ

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & ; \ x \geq 0, \ \theta > 0 \\ 0 & ; \ ow \end{cases}$$

Find MLE $(\hat{\theta})$ the parameter θ and compute variance of $(\hat{\theta})$

Solution

Given probability function is

$$f(x, \theta) = \theta e^{-\theta x} \ ; \ x \geq 0, \ \theta > 0$$

Now, the likelihood function is

$$L(\theta) = \theta^n e^{-\theta \Sigma x_i}$$

$$\log L = n\log\theta - \theta\Sigma x_i$$

$$\log L = n\log\theta - n\theta\bar{x}$$

For MLE

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} - n\bar{x}$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = -\frac{n}{\theta^2} < 0$$

For MLE $\frac{\partial \log L}{\partial \theta} = 0$

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

$$\boxed{\hat{\theta} = \frac{1}{\bar{x}}}$$

This is required MLE of θ

Based on given r.s are

$n = 5$

$x_1 = 0.5$, $x_2 = 1.3$, $x_3 = 0.85$, $x_4 = 2.1$ & $x_5 = 0.40$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$= \frac{0.5 + 1.3 + 0.85 + 2.1 + 0.40}{5}$$

$$= \frac{5.15}{5} = 1.03$$

$$\hat{\theta} = \frac{1}{1.03} = 0.970$$

Now, variance is

$$V(\hat{\theta}) = \frac{1}{I(\theta)}$$

$$\boxed{I(\theta) = E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)} \quad \text{or} \quad \boxed{I(\theta) = E\left(\frac{\partial}{\partial\theta}\log L\right)^2}$$

$$= -E\left(\frac{-n}{\theta^2}\right) = \frac{n}{\theta^2}$$

$$V(\hat{\theta}) = \frac{\theta^2}{n}$$

Now estimated variance

$$V(\hat{\theta}) = \frac{(0.97)^2}{5}$$

$$= 0.188$$

Question

Let $X \sim U(0,\theta)$, Find MLE of θ

If $x_1 = 2.1$, $x_2 = 0.75$, $x_3 = 1.4$, $x_4 = 0.85$, $x_5 = 0.30$ and the r.s from above population then compute the numerical value of $\hat{\theta}$

Solution

Since $X \sim U(0,\theta)$

$$\Rightarrow f(x,\theta) = \frac{1}{\theta}; \quad 0 < x < \theta$$

In this case we know that $\max(x_i\text{'s})$ will be the MLE of θ

$$\Rightarrow \hat{\theta} = \max(x_{(i)})$$

The given value of r.s are

$x_1 = 2.1$, $x_2 = 0.75$, $x_3 = 1.4$, $x_4 = 0.85$, $x_5 = 0.30$

Now ordered form of the given r.s is

$$x_5 < x_2 < x_4 < x_3 < x_1$$

$\min(x_i's) = x_{(5)}$

$\max(x_i's) = x_{(1)}$

$= 2.1$

$\boxed{\hat{\theta} = 2.1}$

MLE and sufficiency.

Let $x_1, x_2 \ldots x_n$ be a r.s of size n from $f(x,\theta)$; $\theta \in \Theta$ and $t = t(x_1, x_2 \ldots x_n)$ be the sufficient statistic for θ

Then by Factorization theorem

$$L(x,\theta) = g(t(x)\theta)\, h(\underset{\sim}{x}) \quad \text{---} ①$$

Taking log both sides, we get

$$\log L(x,\theta) = \log g(t,\theta) + \log h(\underset{\sim}{x})$$

Now differentiating the above expression wrt θ, we have

$$\frac{d}{d\theta} \log L(x,\theta) = \frac{d}{d\theta} \log g(t,\theta) + \cancel{\frac{d}{d\theta} \log h(\underset{\sim}{x})}^{\,0}$$

$$= \frac{d}{d\theta} \log g(t,\theta)$$

For MLE

$$\frac{d}{d\theta} \log L(x,\theta) = 0$$

$$\Rightarrow \left.\frac{d}{d\theta} \log g(t,\theta)\right|_{\theta=\hat{\theta}} = 0 \quad \text{———} ②$$

$$\Rightarrow \hat{\theta} = t(\underset{\sim}{x})$$

= function of sufficient statistic alone

If for any parametric population sufficient statistic t exist then the MLE of the same parametric family will be the function of sufficient statistic

Assumptions and properties of MLE

The maximum Likelihood Estimation for estimating the unknown parameter of the parametric family poses several optimum theory based on few assumptions

1) The 1^{st} and 2^{nd} derivative of the logarithmic of the likelihood function

i.e $\frac{\partial}{\partial\theta}\log L$ and $\frac{\partial^2}{\partial\theta^2}\log L$ should exist and continuous

function of θ in range R including the true value of the Parameter for every θ in R

$$\left|\frac{\partial}{\partial\theta}\log L\right| < F_1(x) \quad \& \quad \left|\frac{\partial^2}{\partial\theta^2}\log L\right| < F_2(x)$$

where $F_1(x)$ and $F_2(x)$ are integrable function on R

2) The 3^{rd} order derivative i.e $\frac{\partial^3}{\partial\theta^3}\log L$ exist such that

$$\left|\frac{\partial^3}{\partial\theta^3}\log L\right| < M(x) \text{ where}$$

$E(M(x)) < k$ a positive quantity
[wrt PDF]

3) For every θ in R

$$E\left(\frac{-\partial^2}{\partial\theta^2}\log L\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\frac{-\partial^2}{\partial\theta^2}\log L\right) L\, dx_1 \ldots dx_n$$

(Crammer-Rao)

is finite and non zero

4) The range of integration should be independent of θ to make differentiation under integration sign valid
But if the range of integration depends on the θ then $f(x,\theta)$ vanishes at the extreme depending on θ

under the above set of assumptions there are following properties that hold good for Maximum Likelihood estimator

* MLE are consistent i.e with probability approaching unity as $n\to\infty$, the likelihood ~~estimation~~ eq^n $\frac{\partial}{\partial\theta}\log L = 0$ has a solution which converges to the true value of θ. This is also

called as Crammer Rao Theorem

* Any consistent solution of the likelihood eqn will provide the maximum of the likelihood with probability tending to unity as the sample size increases ~~independently~~ indefinately

This statement is called Hazoor Bazar's Theorem

* A consistent solution of the likelihood equation is asymptotic normally distributed about the true value of θ

i.e $\hat{\theta} \sim AN(\theta, \frac{1}{I(\theta)})$ where $I(\theta)$ is known as Fisher's information on θ supplied by sample i.e $x_1, x_2 --- x_n$

* If a sufficient estimator exist it is the function of MLE
* If the MLE exist then it is most efficient estimator
* MLE are generally biased
* MLE are not unique in general
* Invariance property hold good for MLE. If $\hat{\theta}$ is MLE for θ and $\phi(\theta)$ is any function of θ then $\phi(\hat{\theta})$ will be MLE for the function of $\phi(\theta)$

<u>Method of Moments</u>

In order to find the point estimate of the parameter for any parametric family, the method of moments is ~~the~~ one of the easiest method for obtaining the point estimates of the parameter

This method was firstly introduced by prof. Karl pearson

This method consist of equating the first few moments of a population to the corresponding moments of sample

Thus getting as many equations as are needed ~~for~~ to solve for unknown parameters of the population

The r^{th} sample moment of a set of observation i.e $x_1, x_2 -- x_n$ is the mean of their r^{th} power and is denoted by m_r' and given by $m_r' = \frac{\sum_{i=1}^{n} x_i^r}{n}$

let r^{th} moment of the population (theoretical) is given by

$$\mu_r' = E(x^r) ; \quad r = 1, 2, 3 --- k \quad 1 \leq r \leq k$$

Let $x_1, x_2 \dots x_n$ be random sample of size n from a population $f(\underset{\sim}{x}, \theta_1, \theta_2 \dots \theta_K)$ then the moment estimators $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_K$ of the parameters $\theta_1, \theta_2 \dots \theta_K$ respectively are the solutions of the simultanous eqn i.e

$$\mu_r'(\theta_1, \theta_2, \theta_3 \dots \theta_K) = m_r'(x_1, x_2 \dots x_K)$$

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

This method of obtaining the estimator of unknown parameter is called as method of moments. This method is also useful to estimate Joint moments

$$\boxed{m'_{(r,s)} = \frac{\sum x_i^r y_i^s}{N}}$$

Properties of Methods of Moments

- ☆ The method of moment is very easy to understand and very easy to compute for any parametric family
- ☆ The moment estimators are consistent
- ☆ The moment estimator is asymptotic normally distributed
- ☆ The moment estimator are not unbiased in general
- ☆ The moment estimator are less efficient
- ☆ The moment estimate and MLE are often identical. But if they do differe the MLE's are usually prefered

Question

Let $X: x_1, x_2 \ldots x_n$ be a r.s of size n from $P(\lambda)$. Then find the moment estimate for the parameter λ

Solution

Given that $X \sim P(\lambda)$

$$\Rightarrow P[X = x] = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{\lfloor x} & ; \quad x = 0, 1, 2 \ldots \infty, \ \lambda > 0 \\ 0 & ; \quad ow \end{cases}$$

The r^{th} moment about origin

$$\mu_r' = E(X^r)$$

if $r = 1$, $\mu_1' = E(X)$

$$\Rightarrow E(X) = \lambda \quad \text{——— (1)}$$

If $x_1, x_2 \ldots x_n$ be a r.s of size n from $P(\lambda)$. Then the sample moment

$$m_r' = \frac{1}{n} \sum_{i=1}^{n} x_i^r$$

If $r = 1$

$$m_1' = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{——— (2)}$$

The moment estimator is obtained by equating (1) & (2)

$$\Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\Rightarrow \boxed{\hat{\lambda} = \bar{x}}$$

Question

Let a r.v $X \sim N(\mu, \sigma^2)$. Find the moment estimator for μ and σ^2

Solution

Given that, the r.v $X \sim N(\mu, \sigma^2)$

$$\Rightarrow f(x, \mu, \sigma^2) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} & ; \quad \mu, x \in R, \ \sigma \in R^+ \\ 0 & ; \quad ow \end{cases}$$

For the above PDF

$$\left.\begin{aligned} E(x) &= \mu \\ E(x^2) &= \sigma^2 + \mu^2 \end{aligned}\right\} \text{ —— (1)}$$

If $x_1, x_2 \dots x_n$ be a r.s of size n from $N(\mu, \sigma^2)$ then the sample moment

$$m_r' = \frac{1}{n} \sum_{i=1}^{n} x_i^r$$

If $r = 1$ $\quad m_1' = \frac{1}{n} \sum_{i=1}^{n} x_i$ —— (2)

If $r = 2$ $\quad m_2' = \frac{1}{n} \sum_{i=1}^{n} x_i^2$

Now by (1) & (2) we get

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

$$\Rightarrow \boxed{\hat{\mu} = \bar{x}}$$

$$\Rightarrow E(x^2) = m_2'$$

$$\Rightarrow \sigma^2 + \mu^2 = m_2'$$

$$\Rightarrow \sigma^2 = m_2' - \mu^2$$

$$\Rightarrow \hat{\sigma}^2 = m_2' - \hat{\mu}^2$$

$$\Rightarrow \hat{\sigma}^2 = m_2' - \bar{x}^2$$

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

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

Question

Find the moment estimator for the parameter α and β for the given prob density function

$$f(x, \alpha, \beta) = \begin{cases} \frac{1}{\beta-\alpha} & ; \quad \alpha < x < \beta \\ 0 & ; \quad ow \end{cases}$$

The given pdf is

$$\left(E(x) = \frac{\alpha+\beta}{2} \right)$$

Now the pop^n moments

$$\mu_r' = E(x^r) = \int_\alpha^\beta x^r f(x, \alpha, \beta)\, dx$$

$$= \int_\alpha^\beta x^r \frac{1}{\beta-\alpha}\, dx$$

$$= \frac{1}{\beta-\alpha} \cdot \left. \frac{x^{r+1}}{r+1} \right|_\alpha^\beta$$

$$= \frac{\beta^{r+1} - \alpha^{r+1}}{(r+1)(\beta-\alpha)}$$

If $r = 1$

$$\mu_1' = E(x) = \frac{\beta+\alpha}{2}$$

$r = 2$

$$\mu_2' = E(x^2) = \frac{\beta^3 - \alpha^3}{3(\beta-\alpha)}$$

$$= \frac{\cancel{(\beta-\alpha)}(\beta^2 + \alpha\beta + \alpha^2)}{3\cancel{(\beta-\alpha)}}$$

$$= \frac{\beta^2 + \alpha\beta + \alpha^2}{3}$$

Now equating population moments to the sample moment, we get,

$\mu_1' = m_1$; where $m_1 = \frac{1}{n} \sum x_i$

$\Rightarrow \frac{\beta + \alpha}{2} = m_1$

$\Rightarrow \alpha + \beta = 2m_1$ —— (1)

again

$\mu_2' = m_2$

$\Rightarrow \alpha^2 + \beta^2 + \alpha\beta = 3m_2$ —— (2)

By (1)

$(\alpha+\beta)^2 = 4m_1^2$

$\alpha^2 + 2\alpha\beta + \beta^2 = 4m_1^2$ —— (3)

by (2)

$3m_2 - \alpha\beta + 2\alpha\beta = 4m_1^2$

$\alpha\beta = 4m_1^2 - 3m_2$ —— (4)

by (1) $\alpha(2m_1 - \alpha) = 4m_1^2 - 3m_2$

$2m_1\alpha - \alpha^2 = 4m_1^2 - 3m_2$

$\Rightarrow \alpha^2 - 2m_1\alpha + 4m_1^2 - 3m_2 = 0$

$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ $a = 1$, $b = -2m_1$, $c = 4m_1^2 - 3m_2$

$\hat{\alpha} = \frac{2m_1 \pm \sqrt{4m_1^2 - 4(4m_1^2 - 3m_2)}}{2}$

$= m_1 \pm \sqrt{m_1^2 - 4m_1^2 + 3m_2}$

$\hat{\alpha} = m_1 \pm \sqrt{3m_2 - 3m_1^2} = m_1 \pm \sqrt{3(m_2 - m_1^2)}$

Now $\hat{\beta} = 2m_1 - \hat{\alpha}$

$$= 2m_1 - \left\{ m_1 \pm \sqrt{3(m_2 - m_1^2)} \right\}$$

Question

For the given pdf. Find moment estimator for θ

$$f(x,\theta) = \begin{cases} \frac{\theta}{x^2} \cdot e^{-\theta/x} & ; \quad x \geq 0, \; \theta > 0 \\ 0 & ; \quad ow \end{cases}$$

Solution

For pdf

$$I = \int \frac{\theta}{x^2} e^{-\theta/x} dx$$

$$\frac{\theta}{x} = P \Rightarrow \frac{-\theta}{x^2} dx = dP$$

$$\int_0^\infty e^{-P} dp = -e^{-P} \Big|_0^\infty = 1$$

Now find

$$\mu_r' = E(x^r)$$

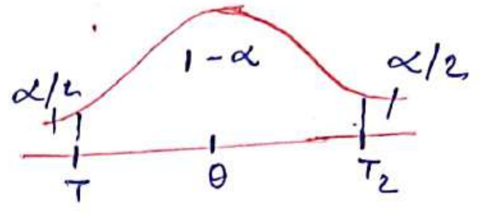
$$= \int_0^\infty x^r \frac{\theta}{x^2} e^{-\theta/x} dx$$

Confidence Interval

Let $x_1, x_2 \dots x_n$ be a r.s from the population $f(x, \theta)$; $\theta \in \Theta$, Let $T_1 = t_1(x_1, x_2 \dots x_n)$ & $T_2 = t_2(x_1, x_2 \dots x_n)$ be the two statistic based on sample value $x_1, x_2 \dots x_n$ satisfying $T_1 < T_2$ for which

$$P[T_1 > \theta] = \alpha_1 \quad \text{——} \quad (1)$$

and $P[T_2 < \theta] = \alpha_2$ —— (2)



Now by combining (1) & (2)

$$P[T_1 < \theta < T_2] = 1-\alpha \quad \text{——} \quad (3)$$

where $\alpha = \alpha_1 + \alpha_2$

The quantity given in R.H.S of eq (3) is independent of the parameter θ. The random interval (T_1, T_2) satisfying (3) is called / known as $100(1-\alpha)\%$ confidence interval. where T_1 & T_2 are the lower and upperlimit of $100(1-\alpha)\%$ CI. The quantity $(1-\alpha)$ is called as confidence coefficient

~~Methods of construction for confider~~

One Sided confidence Interval

Let $x_1, x_2 \dots x_n$ be a r.s from the population $f(x, \theta)$; $\theta \in \Theta$ i.e $x_1, x_2 \dots x_n \overset{iid}{\sim} f(\underline{x}, \theta) \ \forall \ \theta \in \Theta$ and $T_1 = t_1(x_1, x_2 \dots x_n)$ be the any statistic for which

$$P[T_1(x_1, x_2 \dots x_n) < \theta] = 1-\alpha \quad \text{——} \quad (1)$$ then $(-\infty, T_1]$

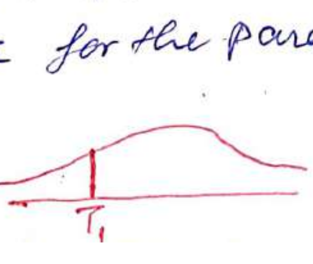
is called as $100(1-\alpha)\%$ one sided lower CI and

Illy $T_2 = t_2(x_1, x_2 \dots x_n)$ be any statistic based on sample values $x_1, x_2 \dots x_n$ for which

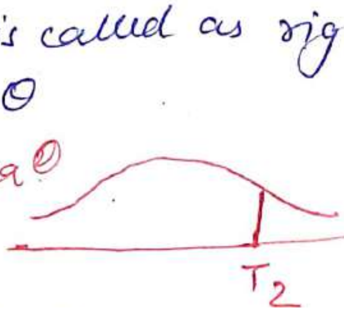
$$P[T_2(x_1, x_2 \dots x_n) > \theta] = 1-\alpha \quad \text{——} \quad (2)$$

Then the random interval $[T_2, \infty)$ is called as right sided $100(1-\alpha)\%$ CI for the parameter θ

eq (1)



eq (2)



Methods for construction of Confidence Interval

These are the following methods for constructing $100(1-\alpha)\%$ CI for the parameter / function of the parameter

① Pivotal quantity Method

② Statistical Method

③ Chebychev's inequality method

Pivotal Quantity

$q(\underset{\sim}{x}, \theta, \sigma^2) \overset{D}{\sim}$ indp of (θ, σ^2)

Let $x_1, x_2 \ldots x_n$ be a rs of size n from $f(x, \theta)$, $\theta \in \Theta$ and $Q = q(x_1, x_2 \ldots x_n, \theta)$ i.e be any function of x_i's and the Parameter θ. if the distribution of Q is independent of Parameter then Q is taken as the pivotal quantity

For eg ⇒ a r.v $X \sim N(\theta, 4)$

$$X \sim N(\theta, 4)$$

$$\bar{x} \sim N\left(\theta, \frac{4}{n}\right)$$

$$\bar{x} - \theta \sim N\left(0, \frac{4}{n}\right) \quad \text{——— (A)}$$

If we define $Q = \bar{x} - \theta$

$= f^n$ of x_is and θ

$\overset{D}{\sim} N\left(0, \frac{4}{n}\right)$

independent of parameter

Hence this $Q = \bar{x} - \theta$ is taken as pivotal quantity

from (A)

$$\underset{\underset{Q}{\sim}}{Z} = \frac{\bar{x} - \theta}{2/\sqrt{n}} \sim N(0, 1)$$

↳ SNV (Standard Normal variance)

↳ independent of parameter

↳ this may be taken as Q i.e pivotal quantity

Hence the $100(1-\alpha)\%$ confidence Interval is computed by

$\Rightarrow P[q_1 < Q(x_1, x_2 \ldots x_n, \theta) < q_2] = 1 - \alpha$

$\Rightarrow \quad q_1 < Q(x_1, x_2 \ldots x_n, \theta) < q_2$

$q_1' < \phi(\theta) < q_2'$

$P[q_1' < \phi(\theta) < q_2'] = 1-\alpha$

the random interval (q_1', q_2') is called as $100(1-\alpha)\%$ CI

Now for the considered problem, the interval is obtained as

$$P[q_1 < Q < q_2] = 1-\alpha$$

$$P\left[q_1 < \frac{\bar{x}-\theta}{2/\sqrt{n}} < q_2\right] = 1-\alpha$$

$$P\left[\frac{2q_1}{\sqrt{n}} < \bar{x}-\theta < \frac{2q_2}{\sqrt{n}}\right] = 1-\alpha$$

$$\cancel{P\left[\frac{2q_1}{\sqrt{n}} \quad P[\bar{x}-2q_2\right.}$$

$$P\left[\underbrace{\bar{x} - \frac{2q_2}{\sqrt{n}}}_{q_1'} < \theta < \underbrace{\bar{x} + \frac{2q_1}{\sqrt{n}}}_{q_2'}\right] = 1-\alpha$$

Hence the $100(1-\alpha)\%$ CI is

$$\left(\bar{x} - \frac{2}{\sqrt{n}}q_1,\ \bar{x} + \frac{2}{\sqrt{n}}q_2\right)$$

Obtain $100(1-\alpha)\%$ confidence interval for the parameter μ where the rv X has $N(\mu, \sigma^2)$ when

a) σ^2 is known

b) σ^2 is not known

<u>Solution</u>

Given that the rv $X \sim N(\mu, \sigma^2)$

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

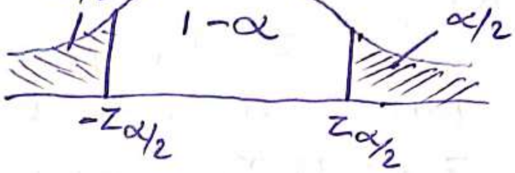
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

we want to construct $100(1-\alpha)\%$ CI for the parameter μ when σ^2 is known

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

$$\bar{x} - \mu \sim N\left(0, \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$



Hence the quantity $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = Q(Z)$

$$P\left[-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

by area property of normal distribution

$$P\left[-Z_{\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq Z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

$$P\left[-\frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} \leq \bar{x} - \mu \leq \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

$$P\left[\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

The $100(1-\alpha)\%$ CI for μ is

$$\left(\bar{x} \mp Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$

width of interval is

$$D = \cancel{\bar{x}} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}} - \cancel{\bar{x}} + \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

$$= 2 \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$$

If $\alpha = 5\%$ then we have 95% CI for the parameter μ

for $\alpha = 5\%$, $Z_{\frac{\alpha}{2}} = 1.96$

for $\alpha = 1\%$, $Z_{\frac{\alpha}{2}} = 2.58$

$\Rightarrow P[-1.96 \leq Z \leq 1.96] = 0.95$

$\parallel$ly $P[-2.58 \leq Z \leq 2.58] = 0.99$

The required interval is

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} , \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

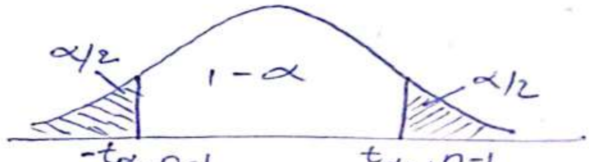
$\parallel$ly the 99% CI for μ

$$\mu \in \left(\bar{x} \mp 2.58 \frac{\sigma}{\sqrt{n}}\right)$$

Case b

when σ^2 is unknown then the pivotal or test statistic is t - test

$$\Rightarrow t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$$



$n-1 \Rightarrow$ Degree of freedom

where

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

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

$s^2 \Rightarrow$ best unbiased estimator for σ^2

$$E(s^2) = \sigma^2$$

$$\boxed{\hat{\sigma}^2 = s^2}$$

Hence the $100(1-\alpha)\%$ CI for μ when σ^2 is unknown

$$P\left[-t_{\alpha/2, n-1} \leq t \leq t_{\frac{\alpha}{2}, n-1}\right] = 1-\alpha$$

$$\Rightarrow P\left[-t_{\frac{\alpha}{2}, n-1} \leq \frac{\bar{x}-\mu}{s/\sqrt{n}} \leq t_{\frac{\alpha}{2}, n-1}\right] = 1-\alpha$$

after simplification we get

$$\Rightarrow P\left[\bar{x} - \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1} \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}\right] = 1-\alpha$$

The required interval is

$$\left(\bar{x} \mp \frac{s}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}\right)$$

Confidence Interval for large sample

It has been proven under certain regularity condition. The first derivative of the logrithm of the likelihood function ~~i.e~~ with respect to parameter θ i.e $\frac{\partial}{\partial\theta}\log L$ is asymptotically normally distributed with mean 0 and variance

$$V\left(\frac{\partial}{\partial\theta}\log L\right) = E\left(\frac{\partial}{\partial\theta}\log L\right)^2$$
$$= E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right)$$

$$\text{i.e } \frac{\partial}{\partial\theta}\log L \sim AN\left(0, E\left(\frac{\partial}{\partial\theta}\log L\right)^2\right)$$

For large n

$$Z = \frac{\frac{\partial}{\partial\theta}\log L}{\sqrt{E\left(\frac{\partial}{\partial\theta}\log L\right)^2}} \sim N(0,1)$$

This quantity Z enables us to construct $100(1-\alpha)\%$ CI for the parameter θ then

$$P\left[|Z| \leq \lambda_\alpha\right] = 1-\alpha$$

$$\Rightarrow P\left[-\lambda_\alpha \leq Z \leq \lambda_\alpha\right] = 1-\alpha \Rightarrow \int_{-\lambda_\alpha}^{\lambda_\alpha} f(z)\,dz = 1-\alpha$$

Question

Construct $100(1-\alpha)\%$ confidence interval for the parameter θ for the given population for large n

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & ; \; x \geq 0 \\ & \;\;\; \theta > 0 \\ 0 & ; \; ow \end{cases}$$

Solution

The ~~pdf~~ given family is

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & ; \; x \geq 0 \\ & \;\;\; \theta > 0 \\ 0 & ; \; ow \end{cases}$$

The likelihood function is

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

$$= \theta^n e^{-\theta \Sigma x_i}$$

Taking logritm both sides, we get

$$\log L = n\log\theta - \theta \sum_{i=1}^{n} x_i$$

$$= n\log\theta - n\theta\bar{x}$$

$$\Rightarrow \frac{\partial}{\partial\theta}\log L = \frac{n}{\theta} - n\bar{x} \quad \text{——— (1)}$$

$$\frac{\partial^2}{\partial\theta^2}\log L = \frac{-n}{\theta^2}$$

$$E\left(-\frac{\partial^2}{\partial\theta^2}\log L\right) = E\left(-\left(-\frac{n}{\theta^2}\right)\right) = \frac{n}{\theta^2}$$

Now for large n, the pivotal is

$$Z = \frac{\frac{n}{\theta} - \bar{x}n}{\sqrt{\frac{n}{\theta^2}}} \sim AN(0,1)$$

$$Z = \frac{n - n\theta\bar{x}}{\sqrt{n}}$$

$$= \sqrt{n}(1-\theta\bar{x})$$

Then $100(1-\alpha)\%$ CI for θ is

$$P\left[|z| \leq Z_{\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\Rightarrow P\left[-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}\right] = 1-\alpha$$

if $\alpha = 5\%$ $\quad Z_{\frac{\alpha}{2}} = 1.96$

$$\Rightarrow P[-1.96 \leq Z \leq 1.96] = 1-\alpha$$

$$\Rightarrow P\left[-1.96 \leq \sqrt{n}(1-\theta\bar{x}) \leq 1.96\right] = 1-\alpha$$

$$\Rightarrow P\left[\frac{-1.96}{\sqrt{n}} \leq 1-\theta\bar{x} \leq \frac{1.96}{\sqrt{n}}\right] = 0.95$$

$$1-\theta\bar{x} \geq -\frac{1.96}{\sqrt{n}}$$

$$1+\frac{1.96}{\sqrt{n}} \geq \theta\bar{x}$$

$$\theta \leq \frac{1}{\bar{x}}\left(1+\frac{1.96}{\sqrt{n}}\right)$$

$$1-\theta\bar{x} \leq \frac{1.96}{\sqrt{n}}$$

$$\theta\bar{x} \geq 1-\frac{1.96}{\sqrt{n}}$$

$$\theta \geq \frac{1}{\bar{x}}\left(1-\frac{1.96}{\sqrt{n}}\right)$$

$$\Rightarrow P\left[\frac{1}{\bar{x}}\left(1-\frac{1.96}{\sqrt{n}}\right) \leq \theta \leq \frac{1}{\bar{x}}\left(1+\frac{1.96}{\sqrt{n}}\right)\right] = 0.95$$

Hence the required 95% CI is

$$\left(1 \mp \frac{1.96}{\sqrt{n}}\right) \frac{1}{\bar{x}}$$

Testing of Hypothesis

Hypothesis: Any statistical statement / assumption / assertion / specification about the population or characteristics of the population is called as hypothesis.

Hypothesis

- Parametric (Specifications about the Parameter of the population)
  - Simple hypothesis
  - Composite Hypothesis
- Non-parametric (Specification about the frequency function)

A ~~Simple~~ hypothesis (H) which completely specify the parameter of the population is called as simple Hypothesis. Otherwise it is composite hypothesis

For eg ⇒ A r.v $X \sim N(\mu, \sigma^2)$ then the hypothesis

$\boxed{\mu=2, \sigma^2=1}$ ← $H: \mu = \mu_0, \sigma^2 = \sigma_0^2$ is the simple hypothesis. However $H: \mu = \mu_0$ or $H: \mu \neq \mu_0, \sigma^2 > 0$ or $H: \mu > \mu_0, \sigma^2 = \sigma_0^2$ are the composite hypothesis

Null Hypothesis

A Hypothesis which is of no differences is called as null hypothesis and it is denoted by H_0

Definition-2

A hypothesis which is to be tested for the possible rejection

Alternative Hypothesis

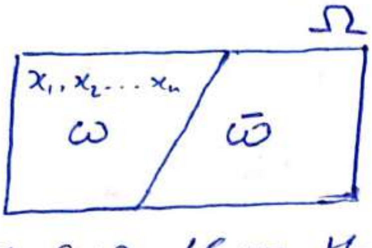
Testing of null Hypothesis H_0 is meaningful if it is tested against certain rival hypothesis. and this rival hypothesis is called as Alternative Hypothesis

Hence, any complement hypothesis to the null hypothesis is called as Alternative hypothesis. usually it is denoted by H_A or H_1

$$H_0 : \mu = 5'4''$$
$$H_1 : \mu > 5'4''$$

Critical region

Let Ω be the sample space and $x_1, x_2 \dots x_n$ be the n sample points such that x_i's $\in \Omega \ \forall \ i = 1, 2, 3 \dots n$. Now dividing Ω into two disjoint regions ω and $\bar{\omega}$ such that $\omega \cap \bar{\omega} = \phi$. If all the sample points fall into the region ω then the null hypothesis H_0 is rejected otherwise we do not reject H_0



If all $x_i \in \omega$ then H_0 is rejected

ω : rejection region

$\bar{\omega}$: Acceptance region

Thus, the critical reason (ω) is that part of sample space where the null hypothesis H_0 is rejected

Error in testing of Hypothesis Problem

Testing of hypothesis is a two action decision problem where the decisions are made ~~based on inf~~ in terms of acceptance or rejection of the hypothesis based on information supplied by sample. The decision drawn in light of sample may not always be true. Hence, we may have the following situation due to acceptance of wrong hypothesis or rejection of true hypothesis

		decision based on sample	
		Reject H_0	Accept H_0
True states of Hypothesis	H_0 is true	Incorrect decision Type-I error	correct decision
	H_0 is False	correct decision	Incorrect decision Type-II error

Steps involved in testing of Hypothesis

* We must have explicit knowledge about the population ~~and~~ under study and the interest of characteristics about which we want to construct the hypothesis

* Setup null and Alternative hypothesis
* Find appropriate test statistic to test that particular hypothesis
* Find critical region based on appropriately choosen level of significance α

* ~~compare~~ Make decision by comparing the test statistic value and the critical value

Reject H_0 when the calculated value of test statistic is greater than the critical value or tabulated value

Type-I error

Size of the test

It is the error arises due to rejection of true hypothesis H_0 when it is true. It is denoted by α

$$\alpha = P[\text{rejecting } H_0 \mid H_0 \text{ is true}]$$
$$= P[\underline{x} \in \omega \mid H_0 \text{ is true}]$$

Type-II error

It is the probability that accepting H_0 when H_1 is true. It is denoted by β

$$\beta = P[\text{accepting } H_0 \mid H_1 \text{ is true}]$$
$$= P[x \in \bar{\omega} \mid H_1 \text{ is true}]$$

Level of significance

It is the supremum value of type-I error. It is also called as size of test and is denoted by α. It is usually taken as 5% and 1% → Permissible error

Power of the test

$1-\beta$ is called as power of test or in other words. It is the probability that the null hypothesis H_0 is rejected where H_1 is true

$$1-\beta = P[x \in \omega \mid H_1]$$

Optimum Test

A test $\phi(x)$ and the corresponding critical region ω is said to be optimum. If it minimizes both types of error i.e α and β simultaneously. But the simultaneous minimization of these two errors are not possible. Hence, we fixed one error i.e type-I and minimize β such that power of test is higher

Most Powerful Test

A critical region ω and the corresponding test for testify simple nulle hypothesis $H_0: \theta = \theta_0$ against simple $H_1: \theta = \theta_1$ said to MP test if

$$P[x \in \omega / H_0] = \alpha \quad \ldots \text{(1)}$$

and

$$P[x \in \omega / H_1] \geq P[x \in \omega_1 / H_1]$$

i.e

$$1 - \beta \geq 1 - \beta_1$$

where ω_1 is the another critical region satisfying (1)

Unformally Most Powerful ~~case~~ Test

A critical region ω and the corresponding test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta$ is said to be UMP test if

$$P[x \in \omega / H_0] = \alpha \quad \text{—— (1)}$$

& $P[x \in \omega / H_1] \geq P[x \in \omega_1 / H_1] \ \forall\ \theta \in \Theta$

where ω_1 is the another critical region of size α

Question

Let us consider a random experiment of tossing of a coin for testing the null hypothesis $H_0 : P = \frac{1}{2}$ against the alternative hypothesis $H_1 : P = \frac{1}{3}$. The hypothesis H_0 is rejected if more than three heads are obtained in five throw. compute type-I, type-II error and power of the test

Solution

Let ~~a~~ X denotes the number of heads obtained in tossing of a coin experiment and given that $n = 5$

$$X \sim B(5, P)$$

$$P[X = x] = \begin{cases} {}^5C_x P^x (1-P)^{5-x} & ; \ 0 < P < 1 \\ & ; \ x = 0,1,2,3,4,5 \\ 0 & ; \ \text{ow} \end{cases}$$

The given hypothesis are

$$H_0 \cdot P = \frac{1}{2}$$

against $H_1 : P = \frac{1}{3}$

under H_0, $X \sim B(5, \frac{1}{2})$

under H_1 , $X \sim B(5, \frac{1}{3})$

The hypothesis H_0 is rejected for

$\omega: x > 3$ this is the critical region

Now types of error are

$$\alpha = P[x \in \omega \mid H_0]$$

$$= P[x > 3 \mid p = 1/2]$$

$$= \sum_{x > 3} P(X = x) \Big| p = \frac{1}{2}$$

$$\alpha = \sum_{x=4}^{5} {}^5C_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{5-x}$$

$$= \sum_{x=4}^{5} {}^5C_x \left(\frac{1}{2}\right)^5$$

$$= \left(\frac{1}{2}\right)^5 \sum_{x=4}^{5} {}^5C_x = \frac{1}{32}\left[{}^5C_4 + {}^5C_5\right]$$

$$= \frac{1}{32}[5+1] = \frac{6}{32} = \frac{3}{16}$$

Now type-II error is

$$\beta = P[x \in \bar{\omega} \mid H_1 \text{ is true}]$$

$\bar{\omega}: x \leq 3$

$$\beta = P[x \leq 3 \mid H_1 : p = \frac{1}{3}]$$

$$= \sum_{x \leq 3} {}^5C_x \left(\frac{1}{3}\right)^x \left(1 - \frac{1}{3}\right)^{5-x}$$

$$= \sum_{x=0}^{3} {}^5C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{5-x}$$

$$= \sum_{x=0}^{3} {}^5C_x \left(\frac{1}{3}\right)^5 (2)^{5-x}$$

$$\beta = \frac{32}{243} \sum_{x=0}^{3} {}^5C_x \left(\frac{1}{2}\right)^x$$

$$= \frac{32}{243} \left\{ {}^5C_0\left(\frac{1}{2}\right)^0 + {}^5C_1\left(\frac{1}{2}\right)^1 + {}^5C_2\left(\frac{1}{2}\right)^2 + {}^5C_3\left(\frac{1}{2}\right)^3 \right\}$$

$$= \frac{32}{243} \left\{ 1 + 5\left(\frac{1}{2}\right) + 10\left(\frac{1}{2}\right)^2 + 10\left(\frac{1}{2}\right)^3 \right.$$

$$= \frac{32}{243} \left\{ 1 + \frac{5}{2} + \frac{10}{4} + \frac{10}{8} \right\}$$

$$= \frac{32}{243} \left\{ \frac{8 + 20 + 20 + 10}{8} \right\}$$

$$= \frac{32}{243} \times \frac{58}{8} = \frac{232}{243}$$

Power of the test

$$1 - \beta = 1 - \frac{232}{243} = \frac{11}{243}$$

$$\left(\frac{1}{3}\right)^x \left(\frac{1}{3}\right)^{5-x} (2)^{5-x}$$

or

$$1 - \beta = P[x \in \omega \mid H_1]$$

$$= P\left[x > 3 \mid P = \frac{1}{3}\right]$$

Question-2.

If $x \geq 1$ be the critical region for the testing of the hypotesis $H_0: \theta = 2$ against $H_1: \theta = 1$ for the given population based on single sample

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & ; x \geq 0, \ \theta > 0 \\ 0 & ; ow \end{cases}$$

Find size and power of test

Solution

The given critical region is

$$\omega : x \geq 1$$
$$\bar{\omega} : x < 1$$

The given hypothesis are,

$$H_0 : \theta = 2 \text{ vs } H_1 : \theta = 1$$

$\Rightarrow$ Size of the test $\alpha = P[x \in \omega \mid H_0]$

$$= P[x \geq 1 \mid \theta = 2]$$

$$\alpha = \int_1^\infty 2 \cdot e^{-2x} dx = 2\int_1^\infty e^{-2x} dx$$

$$= 2\left\{\frac{-e^{-2x}}{2}\right\}_1^\infty$$

$$= -\{e^{-\infty} - e^{-2}\}$$

$$= e^{-2}$$

Power of the test

$$1 - \beta = P[x \in \omega \mid H_1]$$

$$= P[x \geq 1 \mid \theta = 1]$$

$$= \int_1^\infty e^{-x} dx$$

$$= e^{-1} = \frac{1}{e}$$

Unbiased Test

A critical region ω and the corresponding test based on ω is said to be unbiased test if the power of the test exceeds the size of the test i.e,

$$\text{Power} \geq \text{size of the test}$$

$$1-\beta \geq \alpha$$

$$P[x \in \omega / H_1] \geq P[x \in \omega / H_0]$$

Neyman - Pearson Lemma

The lemma provides the most powerful critical region for the testing of the simple null hypothesis against simple alternative hypothesis

Statement :

Let ω be the critical region and k be the any constant such that $k > 0$ such that

$$\omega : \begin{cases} x \in S : \dfrac{f(x, \theta_1)}{f(x, \theta_0)} \geq k \\ x \in S \quad \dfrac{L_1}{L_0} \geq k \end{cases} \qquad \bar{\omega} : \left\{ x \in S : \dfrac{L_1}{L_0} < k \right.$$

where θ_0 and θ_1 are the specified value under the hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ and L_0, L_1 are the likelihood f^n under H_0 & H_1. Then this provides Base critical Region (BCR) or MP critical region of size α

Question

Let a rv X ~~has the prob. distribution~~ is observed from the population

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & ; x \geq 0, \theta > 0 \\ 0 & ; ow \end{cases}$$

for the testing of the hypothesis $H_0 : \theta = 2$ against $H_1 : \theta = 1$. Find the MP test or critical region for the testing of H_0 vs H_1

Solution

The given hypothesis are

$$H_0 : \theta = 2$$
$$H_1 : \theta = 1$$

and the distribution of rv X is

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & ; x \geq 0 \\ & \ \ \theta > 0 \\ 0 & ; ow \end{cases}$$

Now, the likelihood function is

$$L(x,\theta) = \theta^n . e^{-\theta \sum_{i=1}^{n} x_i} \quad \text{—— (1)}$$

Then LF under H_0,

$$L(x,\theta_0) = \theta_0^n e^{-\theta_0 \sum_{i=1}^{n} x_i} \quad \text{—— (2)}$$

and under H_1 is

$$L(x,\theta_1) = \theta_1^n e^{-\theta_1 \sum_{i=1}^{n} x_i} \quad \text{—— (3)}$$

Now, the ratio of LF is

$$\frac{L_1}{L_0} = \frac{\theta_1^n e^{-\theta_1 \sum_{i=1}^{n} x_i}}{\theta_0^n e^{-\theta_0 \sum_{i=1}^{n} x_i}}$$

$$\Rightarrow \frac{L_1}{L_0} = \left(\frac{\theta_1}{\theta_0}\right)^n e^{-(\theta_1 - \theta_0) \sum_{i=1}^{n} x_i}$$

Then by N-P lemma

$$w : \frac{L_1}{L_0} \geq k \quad \text{for any } k > 0$$

$$\Rightarrow \left(\frac{\theta_1}{\theta_0}\right)^n e^{-(\theta_1 - \theta_0) \sum_{i=1}^{n} x_i} \geq k$$

$$\Rightarrow e^{-(\theta_1 - \theta_0) \sum x_i} \geq \left(\frac{\theta_0}{\theta_1}\right)^n . k$$

$$\Rightarrow e^{-(\theta_1 - \theta_0) \sum x_i} \geq k_1 \qquad \left(\left(\frac{\theta_0}{\theta_1}\right)^n . k = k_1 \right)$$

$$-(\theta_1-\theta_0)\sum x_i \geq \log k_1$$

$$\Rightarrow \quad (\theta_0-\theta_1)\sum_{i=1}^{n} x_i \geq \log k_1$$

$\Rightarrow$ If $\theta_0 > \theta_1$

$$(\theta_0-\theta_1)\sum x_i \geq \log k_1$$

$$\sum x_i \geq \frac{\log k_1}{(\theta_0-\theta_1)}$$

$$\boxed{\sum x_i \geq k_2}$$

If $\theta_0 < \theta_1$

$$\boxed{\sum x_i \leq k_2}$$

and the value of k_2 is obtained with the help of type-I error

$$P[x \in \omega \mid H_0] = \alpha$$

$$\Rightarrow P\left[\sum x_i > k_2 \mid H_0 : \theta = \theta_0\right] = \alpha$$

$$P[Z > k_2 \mid H_0 : \theta = \theta_0] = \alpha \qquad \left[Z = \sum x_i\right]$$

$$\int_{k_2}^{\infty} f(z)\,dz \Big|_{\theta=\theta_0} = \alpha$$

$\omega : \sum x_i \geq k$

$k = 1$

$\omega : \sum x_i \geq 1$

if $n = 1$

$\omega : x \geq 1$

Likelihood Ratio Test (L-R Test)

Neyman-Pearson lemma provides an elegant way for testing of the hypothesis simple vs simple alternative hypothesis. But this lemma is not applicable when the nature of hypothesis is not simple vs simple. Thus Neyman pearson proposed another generalized class of testing procedure named as likelihood ratio test (L-R test) for testing the simple or composite vs simple or composite alternative hypothesis. Hence N-P lemma will be the particular case of L-R test. This test is based on maximum likelihood estimator

Construction of L-R test

Let $x_1, x_2 --- x_n$ be a random sample from the population $f(\underset{\sim}{x}, \theta_1, \theta_2 --- \theta_K)$; $\theta_i \in \Theta$; $i = 1, 2 \ldots K$. Θ is called as space of totality. The likelihood function based on $x_1, x_2 --- x_n$

$$L(\underset{\sim}{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta_1, \theta_2 --- \theta_K) \quad \text{(1)}$$

Since L-R test is based on MLE. The MLE of the parameter θ_i ; $i = 1, 2, \ldots K$ is obtained by solving the following normal eq^n

$$\frac{dL}{d\theta_i} = 0 \quad \forall\ i = 1, 2 \ldots k$$

under the condition that

$$\frac{d^2L}{d\theta_i^2} < 0$$

Then $\hat{\theta}_i$ will be the MLE of the parameter

We want to test the hypothesis

$$H_0 : \theta_1, \theta_2 --- \theta_K \in \Theta_0$$

$$\text{vs} \quad H_1 : \theta_1, \theta_2 \ldots \theta_K \in \Theta - \Theta_0$$

So the criterion for testing of null hypothesis i.e H_0 against H_1 is the ratio of the two maximized likelihood function under subspace and whole space Θ

$$\lambda = \lambda(x_1, x_2 \dots x_n) = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta}_1)} = \frac{L_0}{L_1}$$

$$= \frac{\sup_{\theta \in \Theta_0} L(\Theta_0)}{\sup_{\theta \in \Theta_1} L(\Theta_1)}$$

λ is a constant which is always less than unity

$$\Theta_0 \leq \Theta$$

$$L(\hat{\Theta}_0) \leq L(\hat{\Theta})$$

$$\frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} \leq 1$$

$$\boxed{\lambda \leq 1}$$

The hypothesis is rejected if

$$\omega : \lambda \leq \lambda_0$$

The value of λ_0 is computed by

$$P[x \in \omega \mid H_0] = \alpha$$

$$\Rightarrow P[\lambda \leq \lambda_0 \mid H_0] = \alpha$$

$$\Rightarrow \int_0^{\lambda_0} g(\lambda) d\lambda \Big|_{H_0} = \alpha$$

Properties of L-R test

① The L-R test is institutive and lead to same result as N-P lemma when the nature of hypothesis is simple vs simple

② This test is based on MLEs

③ In L-R test the probability of type-I error is controlled by suitably chosen cutoff point λ_0

4) under certain conditions, the ratio λ has asymptotically Chi-square distribution

$$-2\log\lambda \sim \chi^2_{(r)}$$

(5) L-R test are UMP test - if it is exist

(6) under certain condition L-R test is consistent

Applications

(1) L-R test provides the test for single mean of normal population

(2) Test for equality of two means from two independent normal pop^n

(3) Test for single variance of normal pop^n

(4) Test for equality of two variances of two normal pop^n

Question

Let $x_1, x_2 \dots x_n$ be a r.s of size n from $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown construct L-R test for testing of $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$

1st application of LR (Test of single mean from $N(\mu, \sigma^2)$)

(II) $X \sim N(\mu_1, \sigma_1^2)$

$Y \sim N(\mu_2, \sigma_2^2)$

$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ are unknown

$$\left.\begin{array}{l} H_0: \mu_1 = \mu_1 \\ H_1: \mu_1 \neq \mu_2 \end{array}\right\}$$

Two tailed test

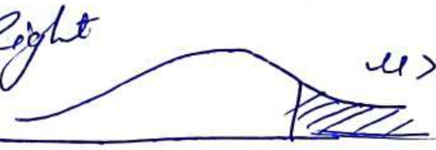
$\mu = \mu_0$

Case-I

$$\sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ (say)}$$

Right

$\mu > \mu_0$



Case-II

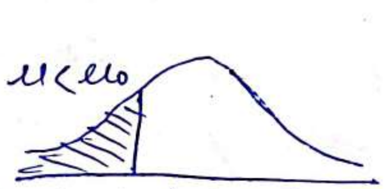
$$\sigma_1^2 \neq \sigma_2^2$$

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

$H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 \neq \sigma_0^2$

$\mu < \mu_0$



left tailed test

$H_1 : \sigma^2 > \sigma_0^2$

$H_1 : \sigma^2 > \sigma_0^2$

(IV) $X \sim N(\mu_1, \sigma^2)$

$Y \sim N(\mu_2, \sigma_2^2)$

$H_0 \sim \sigma_1^2 = \sigma_2^2$

$H_1 \sim \sigma_1^2 \neq \sigma_2^2$

Solution

Given that

$$x_1, x_2 \dots x_n \sim N(\mu, \sigma^2)$$

$$\Rightarrow f(x_i, \mu, \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2} \frac{(x_i - \mu)^2}{\sigma^2}} & ; \mu, x \in \mathbb{R} \\ & \sigma \in \mathbb{R}^+ \\ 0 & ; \text{ow} \end{cases}$$

we want to test the hypothesis

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

where both the parameters μ and σ^2 are unknown

The parameter spaces are

$$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \ 0 < \sigma^2 < \infty\}$$

$$\Theta_0 = \{(\mu, \sigma^2) : \mu = \mu_0, \ 0 < \sigma^2 < \infty\}$$

Now the likelihood function is

$$L(\underline{x}, \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2} \frac{(x_i - \mu)^2}{\sigma^2}}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-1/2 \sum \frac{(x_i - \mu)^2}{\sigma^2}}$$

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

For MLE of μ and σ^2

$$\left.\begin{array}{ll} \frac{d}{d\mu}\log L = 0 & ; \frac{d^2}{d\mu^2}\log L < 0 \\ \& \ \frac{d}{d\sigma^2}\log L = 0 & ; \frac{d^2}{d\sigma^4}\log L < 0 \end{array}\right\}$$

$$\Rightarrow \frac{d}{d\mu}\log L = 0 - 0 + \frac{1}{2\sigma^2}\Sigma(x_i - \mu) = 0$$

$$\Rightarrow \Sigma(x_i - \mu) = 0$$

$$\Rightarrow \boxed{\hat{\mu} = \bar{x}}$$

check $\frac{d^2}{d\mu^2}\log L < 0$

Now, again

$$\frac{d}{d\sigma^2}\log L = 0 - \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$= \frac{n}{2\sigma^2} = \frac{1}{2\sigma^4}\Sigma(x_i - \mu)^2$$

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

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

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

Hence under Θ

$\hat{\mu} = \bar{x}$

$\hat{\sigma}^2 = s^2$

Then $$L(\hat{\Theta}) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{n/2} e^{-\frac{1}{2}\sum_i \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^2}}$$

$$= \left(\frac{1}{2\pi s^2}\right)^{n/2} e^{-\frac{1}{2}\frac{\Sigma(x_i - \bar{x})^2}{s^2}}$$

$$L(\hat{\Theta}) = \left(\frac{1}{2\pi s^2}\right)^{n/2} e^{-n/2} \quad \text{——— (1)}$$

Now, under Θ_0, the LF is ; $H_0 : \mu = \mu_0$

$$L(\Theta_0) = \left(\frac{1}{2\pi\sigma^2}\right)\sigma^{-\frac{1}{2}\sum\frac{(x_i-\mu_0)^2}{\sigma^2}}$$

The MLE of σ^2 is computed by proceeding on same lines we have done earlier

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum(x_i-\mu_0)^2$$

Then,

$$L(\hat{\Theta}_0) = \left(\frac{1}{2\pi\hat{\sigma}_0^2}\right)^{n/2} \cdot e^{-\frac{1}{2}\sum\frac{(x_i-\mu_0)^2}{\hat{\sigma}_0^2}}$$

$$= \left(\frac{1}{2\pi\hat{\sigma}_0^2}\right)^{n/2} \cdot e^{-n/2} \quad \text{——— (2)}$$

Thus, the ratio λ is

$$\lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\left(\frac{1}{2\pi\hat{\sigma}_0^2}\right)^{n/2}\cancel{e^{-n/2}}}{\left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{n/2}\cancel{e^{-n/2}}}$$

$$= \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}$$

$$= \left\{\frac{s^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2}\right\}^{n/2}$$

$$\sum(x_i-\mu_0)^2 = \sum_i(x_i-\bar{x}+\bar{x}-\mu_0)^2$$

$$= \sum_i(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2$$

$$= ns^2 + n(\bar{x}-\mu_0)^2$$

$$\lambda = \left(\frac{ns^2}{ns^2+n(\bar{x}-\mu_0)^2}\right)^{n/2}$$

$$\lambda = \left[\frac{1}{1+\left(\frac{\bar{x}-\mu_0}{s}\right)^2}\right]^{n/2}$$

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

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

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

$$= S^2$$

$$\boxed{s^2 = \left(\frac{n-1}{n}\right)S^2}$$

$$\lambda = \left[\frac{1}{1+\frac{(\bar{x}-\mu_0)^2}{\frac{(n-1)}{n}S^2}}\right]^{n/2}$$

$$= \left[\frac{1}{1+\frac{1}{(n-1)}\left(\frac{\bar{x}-\mu_0}{S/\sqrt{n}}\right)^2}\right]^{n/2}$$

we know that

$$t = \frac{\bar{x}-\mu_0}{S/\sqrt{n}} \sim t_{(n-1)}$$

$\hookrightarrow$ t - statistic

$$\Rightarrow \lambda = \left(1+\frac{t^2}{n-1}\right)^{-n/2}$$

$= f(t^2)$ monotone f^n of t

$\Rightarrow$ Reject H_0:

$$\omega : \lambda \leq \lambda_0$$

$$\Rightarrow \left(1+\frac{t^2}{n-1}\right)^{-n/2} \leq \lambda_0$$

$$\Rightarrow 1+\frac{t^2}{n-1} \geq \lambda_0^{-2/n}$$

$t^2 \leq \lambda_1 \qquad \lambda_1 = (\lambda_0^{-2/n} - 1)(n-1)$

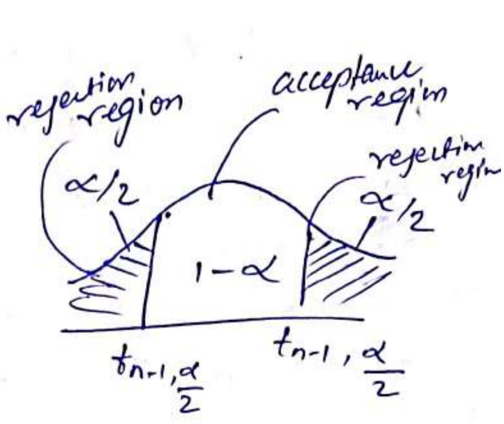
$\boxed{|t| \leq \lambda_2}$

$\omega : |t| \leq \lambda_2$

The value of λ_2 is computed by

$P[x \in \omega \mid H_0] = \alpha$

$P[|t| < \lambda_2 \mid H_0] = \alpha$



Test : two tailed test

Decision Theory

All the problem of statistical Inference may be viewed as problem of decision making in face of uncertainity. Thus, the Problem of point estimation is the problem which values in the parameter space is the true value of the parameter θ.

The problem of hypothesis testing, is to decide whether the null hypothesis is to be taken as true or false and the problem of interval estimation is taken to decide which interval is more accurate and containing the true value of the parameter. In all the above estimation procedures, decision is taken in light of observed sample information i.e $x_1, x_2 --- x_n$

All the above discussed problems are different to each other The problem of decision theory is totally different from the above methods. This problem is based on the measurement of uncertainity in face of decision.

For eg:) A manufacturer of medicine has 5 methods to produce medicines. you would like to select method that would lead to maximum net profit. Hence, in this case you'll make 5 possible decisions based on each of the above 5 methods. In solving such problems you would of course consider a sample product manufactured by each of the methods

Thus, this problem doesn't fall in none of the above methods.
Therefore, making decision of the problems in presence of uncertainity or randomness is known as decision theoretic approach

Element of decision theory

Decision theory problems includes 3 basic elements

① A non-empty set Θ i.e set of all possible states of nature and known as parameter space

② A non-empty set $\mathcal{A}$ (script A) i.e set of all possible states of actions available to staticians

③ Earlier by function $L(\theta, a)$ i.e loss function defined over $\Theta \times \mathcal{A}$. Sometimes denoted by $L(\theta, d)$

Loss function

The most important element of decision theory is loss function. It is a non-negative real value function of θ and a defined over cartesian product of $\Theta \times \mathcal{A}$ and it is denoted by $L(\theta, a)$

$$Loss = L(\theta, a)$$

This function depends over θ & a only

Types of loss function

① Squared error loss f^n

$$L(\theta, d) = (\theta - d)^2 \rightarrow \text{posterior mean} = \hat{d}$$

② Absolute loss function

$$L(\theta, d) = |\theta - d| \Rightarrow d = \text{posterior median}$$

③ Zero one loss f^n

$$L(d, \theta) = \frac{1}{n}\sum_{i=1}^{n} \delta_{d \neq \theta} \xrightarrow{} 0 \text{ with } \delta_{d=\theta} \to 1 \Rightarrow d = \text{posterior mode}$$

④ Linex loss function (Linear exponential loss function)

⑤ General entropy loss function

Decision rule

Let x be a rv with pdf $x \sim f(x, \theta) : \theta \in \Theta$

Then due to this random variable x we have a space of outcome of random variable $\mathcal{X}$ i.e $x \in \mathcal{X}$

Thus. after taking the value of x, we define a function $d(x)$ which maps from sample space $\mathcal{X}$ in action space $\mathcal{A}$ is called as decision rule. Hence, decision ~~rup~~ rule is a mapping from sample space to the action space

$$d : \mathcal{X} \longrightarrow \mathcal{A}$$

Such that

$$d(x) = a \quad \forall \; a \in \mathcal{A}$$

Thus, our loss func^n is converted into $L(\theta, d)$, where d is the decision available to the Staticians.

Risk function

In decision theory, our aim is to minimize loss function i.e $L(\theta, d(x))$ but unfortunately the minimization of $L(\theta, d(x))$ is not logical because $d(x)$ is a random quantity and minimization of a random variable have no sense. Thus we go for risk function i.e expected loss function over whole sample space. Thus, risk function is defined as the expected value of loss function and is denoted by $R(\theta, d)$

$$Risk = E(L(\theta, d(x)))$$

$$= \int_{\mathcal{X}} L(\theta, d(x)) f(x, \theta) dx$$

Non-randomized decision rule

Any function $d(x)$ which map from sample space to action space is called as non-randomized decision rule provided the risk function of $d(x)$ exist at finite $\forall \; \theta \in \Theta$

Baye's Principle

The Baye's principle involves the notion of distribution of the Parameter θ on parameter space Θ called as prior disbution say $g(\theta)$.

Two things require for $g(\theta)$

① we must able to speak about Baye's risk of the decision rule δ wrt prior distⁿ $g(\theta)$

② we need to have joint distribution of $g(\theta)$ & X, and the conditional distribution $g(\theta)|X$, called as posterior distⁿ

$$X \sim f(x, \theta) : \theta \in \Theta$$ → pdf

θ is a parameter

θ is unknown but not fixed

↳ treated as a rv

↳ uncertainity associated with the parameter θ will be quantified in terms of distribution called as prior distribution and is denoted by $g(\theta)$

$$\theta \sim g(\theta) \rightarrow \text{Prior distribution}$$

$$X|\theta \sim f(x,\theta)$$
$$\theta \sim g(\theta)$$

The LF is

~~Joint dist~~ $$L(x|\theta) = \prod_{i=1}^{n} f(x,\theta)$$

Joint dist

$$h(x,\theta) = L(x|\theta)\cdot g(\theta)$$

The posterior distⁿ

$$\pi(\theta|\underset{\sim}{x}) = \frac{h(x,\theta)}{m(x)}$$

where

$$m(x) = \begin{cases} \sum_{\theta} h(x,\theta) & \text{for discrete} \\ \int_{\theta} h(x,\theta) & \text{for continuous} \end{cases}$$

$$\pi(\theta|\underset{\sim}{x}) = \frac{L(x|\theta)\cdot g(\theta)}{\int L(x,\theta)\cdot g(\theta)\,d\theta} = K\cdot \underbrace{L(x|\theta)\cdot g(\theta)}_{\text{Prior} \times \text{likelihood}} \quad \text{where } K = \frac{1}{\int L(x,\theta)\cdot g(\theta)\,d\theta}$$

$\pi(\theta|\underset{\sim}{x})$ = Posterior distⁿ

This is the distribution which contributes the information supplied by sample and distribution of parameter. It is the updated distⁿ of θ obtained by Bayes theorem

The Bayes estimate under squared error loss fⁿ is the Posterior mean

$$E(\theta|\underset{\sim}{x}) = \int \theta . \pi(x|\underset{\sim}{\theta})\, d\theta$$

$$\hat{\theta} = \text{Bayes estimate}$$

Types of trial

① proper prior

A prior $g(\theta)$ is said to be proper

$$\text{if } \int_{\theta} g(\theta) d\theta = 1$$

② Improper prior

$$\int_{\theta} g(\theta) d\theta \neq 1$$

③ Informative and non-informative prior

④ conjugate prior

Question

Let a rv X has the distribution

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & ; x \geq 0, \theta \geq 0 \\ 0 & ; ow \end{cases}$$

and the parameter θ has uniform distribution

$$g(\theta) = \begin{cases} 1 & ; 0 < \theta < 1 \\ 0 & ; ow \end{cases}$$

Find Bayes estimate of the parameter θ under squared error loss fⁿ

Solution

$X \sim f(x,\theta)$

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x} & ; x \geq 0, \theta > 0 \\ 0 & ; ow \end{cases}$$

$$g(\theta) = \begin{cases} 1 & ; 0 < \theta < 1 \\ 0 & ; ow \end{cases}$$

The likelihood function is

$$L(x,\theta) = \theta^n e^{-\theta \Sigma x_i}$$

$$\pi(\theta | \underset{\sim}{x}) = \frac{L(x|\theta) \cdot g(\theta)}{\int L(x,\theta) \cdot g(\theta) \, d\theta}$$

$$\pi(\theta, \underset{\sim}{x}) = \frac{\theta^n e^{-\theta \Sigma x_i} \cdot 1}{\int_{\theta=0}^{\infty} \theta^n e^{-\theta \Sigma x_i} d\theta}$$

$$I = \int_0^{\infty} \theta^n e^{-\theta \Sigma x_i} \, d\theta$$

$\theta \Sigma x_i = P$

$\cancel{\frac{d\theta}{dP}}$ $d\theta = \frac{dP}{\Sigma x_i}$

Range

$\theta \to 0 \Rightarrow P \to 0$

$\theta \to \infty \Rightarrow P \to \infty$

$$I = \int_0^{\infty} \frac{P}{\Sigma x_i} e^{-P} \frac{dP}{\Sigma x_i}$$

$$= \frac{1}{(\Sigma x_i)^{n+1}} \int_0^{\infty} P^{n+1-1} e^{P} dP$$

$$= \left(\frac{1}{\Sigma x_i}\right)^{n+1} \cdot \overline{|n+1}$$

$$= \frac{\overline{|n+1}}{(\Sigma x_i)^{n+1}}$$

$$\pi(\theta|\underset{\sim}{x}) = \frac{\theta e^{-\theta \Sigma x_i}}{\frac{\Gamma n+1}{(\Sigma x_i)^{n+1}}}$$

$$\pi(\theta|\underset{\sim}{x}) = \frac{(\Sigma x_i)^{n+1} \theta^n e^{-\theta \Sigma x_i}}{\Gamma n+1}$$

Now

$$\hat{\theta}_{Bayes} = E(\theta/\underset{\sim}{x}) = \int_0^\infty \theta \cdot \pi(\theta/\underset{\sim}{x}) d\theta$$

$$= \int_0^\infty \frac{\theta(\Sigma x_i)^{n+1} \cdot \theta^n e^{-\theta \Sigma x_i}}{\Gamma n+1} d\theta$$

$$\hat{\theta}_B = \frac{(\Sigma x_i)^{n+1}}{\Gamma n+1} \int_0^\infty \theta^{n+1} e^{-\theta \Sigma x_i} d\theta$$

$$\theta \Sigma x_i = z$$

$$d\theta = \frac{dz}{\Sigma x_i}$$

$$= \frac{(\Sigma x_i)^{n+1}}{\Gamma n+1} \int_0^\infty \left(\frac{z}{\Sigma x_i}\right)^{n+1} e^{-z} \frac{dz}{\Sigma x_i}$$

$$= \frac{1}{(\Sigma x_i)\Gamma n+1} \cdot \Gamma n+2$$

$$= \frac{\Gamma n+2}{(\Sigma x_i)\Gamma n+1}$$

$$\hat{\theta}_B = \frac{(n+1)\cancel{\Gamma n+1}}{(\Sigma x_i)\cancel{\Gamma n+1}}$$

$$\boxed{\hat{\theta}_B = \frac{n+1}{n\bar{x}}}$$

Question

Let a rv X has prob f^n

$$f(x,p) = \begin{cases} p^x (1-p)^{1-x} & ; \ x = 0 \\ & \ \ 0 < p < 1 \\ 0 & ; \ ow \end{cases}$$

and
$$g(p) = \begin{cases} 1 & ; \ 0 < p < 1 \\ 0 & ; \ ow \end{cases}$$

find Bayes estimator for the parameter p under SELF