Module - I

Theory of Estimation

01: Point Estimation

* The main objective of statistical analysis is to draw valid conclusion about (P) (population) on the basis of a sample drawn from the (P).
* Statistical inference is the process of making inferances about the unknown aspects of distri... of the (P) based on samples taken from it.
* The unknown aspects may be the form of distri... / the values of paramtrs involved in it.
* Statistical inference is indeactive natre, bcz the conclusions are drawn from the sample about the unknown aspects of (P).
* Statistical inference has been dichotomiced into 2 areas →

① Estimation of paramtr.

② Test of hypothesis. = It is a type of statistical decission prbm. It is a rule / procedure whether to accept / reject to the hypothesis on the basis of sample values obtained.

* The technique of coming to conclusion regarding the values of the unknown paramtrs based on the info provided by a sample → prbm of estimation

This estimation can be made into 2 ways

1) point estimation (fixed amount / value, eg→200) single value

2) Interval estimation (ranges → 10-100)

⇒ **Point estimation** = (P.E)

* If from the observation in a sample a single value is calculated as an estimate of the unknown paramtr.

• The procedure is → P.E.

* And we refer to the value of the statistic as a point estimate.

* The term estimated represents a rule / a method of estimating the (P) paramtr & the estimate represents the value produced by the estimator.

→ properties →

* There are 4 criteria commonly used for finding a good estimator, They are,

1) Unbiasedness estimator
2) Consistency
3) Efficiency
4) Sufficiency.

① **Unbiasedness** ⇒

It is an statistic that has an expected value equal to the unknown true value of the (P) paramtr being estimated.

* An estimator not having this (pro) is said to be biased.

* Let 'X' be r.v having the pdf $f(x, \theta)$ where θ may be unknown. Let $x_1, x_2 \dots x_n$ be a random sample taken from the (P) represented by X. Let $t_n = t(x_1, x_2, \dots x_n)$ be an estimator of paramtr 'θ'.

t_n → estimator
$E(t_n)$ → expectation of t_n

* If $\boxed{E(t_n) = \theta}$ (expectation → E) for every n then the estimator

t_n → unbiased estimator.

Q
1) Let $x_1, x_2 \dots x_n$ be a random sample drawn from a gvn (P) with mean μ and variance σ^2, S.T the sample mean $\bar{x}$ is an unbiased estimator of (P) mean μ.

A) To prove $E(\bar{x}) = \mu$.

we know that $\boxed{\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i}$

we need expectation, so

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

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

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

x → with mean μ
i.e. $E(x) = \mu$

$$= \frac{1}{n}\left[\underbrace{\mu + \mu + \dots + \mu}_{n \text{ times } \mu}\right]$$

$$= \frac{1}{\cancel{n}} \times \left[\cancel{n}\mu\right]$$

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

Hence sample mean is an unbiased estimator of (P) mean.

2) Let $x_1, x_2, \dots x_n$ is an random sample from a normal distri.. $N(\mu, 1)$ ($N(\mu, \sigma)$), S.T $t = \frac{1}{n} \sum_{i=1}^{n} x_i^2$ is an unbiased estimator of $\mu^2 + 1$.?

A) $N(\mu, \sigma)$

$N(\mu, 1) \to q$; $\sigma = 1$ (variance)

$\boxed{E(x) = \mu \quad E(x)^2 = \mu^2 + 1}$

gvn

$E(x_i) = \mu$

$V(x_i) = 1$

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

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

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

$E(x)^2 - \mu^2 - 1 = 0$

$\therefore E(x_i)^2 = \mu^2 + 1$

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

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

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

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

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

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

$\therefore$ t is an unbiased estimate μ^2+1.

② **Consistency** =

The estimator $t_n = t(x_1, x_2 \dots x_n)$ of paramtr $\theta \to$ consistnt

If t_n converges to θ in (prob). (ie) for $\varepsilon > 0$, $\lim_{n\to\infty} P(|t_n - \theta| \le \varepsilon) = 1$ or

$\lim_{n\to\infty} P(|t_n - \theta| \ge \varepsilon) = 0$.

The estimators satisfying above condin. are $\to$ weakely consistnt estimators

Theorem $\to$

An estimator t_n is such that $E(t_n) = \theta_n \xrightarrow{\text{cgsto}} \theta$ & $V(t_n) \xrightarrow[\text{cgsto}]{} 0$ as $n \to \infty$.

The estimator t_n is said to be consistnt for θ.

Pf/

by chebycheff's inequality $P(x - E(x) \ge k) \le \frac{V(x)}{k^2}$ for any gvn no.. howevr small $\varepsilon > 0$, we hve

$$\boxed{\begin{gathered} P(|t_n - \theta_n| \ge \varepsilon) \le \frac{V(t_n)}{\varepsilon^2} \quad \text{or} \\ P(|t_n - \theta_n| < \varepsilon) > 1 - \frac{V(t_n)}{\varepsilon^2} \end{gathered}}$$

Thus we can write $|t_n - \theta_n| = |t_n - \theta - (\theta_n - \theta)|$

$> |t_n - \theta| - |\theta_n - \theta|$.

$P(|t_n - \theta| - |\theta_n - \theta| < \varepsilon) > 1 - \frac{V(t_n)}{\varepsilon^2}$

$P(|t_n - \theta| < |\theta_n - \theta| + \varepsilon) > 1 - \frac{V(t)}{\varepsilon^2}$.

Since $\theta_n \to \theta$, $|\theta_n - \theta| = \varepsilon_1$

hence $P(|t_n - \theta| < \varepsilon_1 + \varepsilon) > 1 - \frac{V(t)}{\varepsilon^2}$

Since $V(t) \to 0$, hence $\frac{V(t_n)}{\varepsilon^2} < \delta$ (delta)

$P(|t_n - \theta| < \varepsilon) > 1 - \delta$

hence t_n is a consistnt estimator $(t_n \to \theta)$

1) If t is consistnt estimator of θ then t^2 is also a consist (e) of θ^2,

A) Since t is c.(e) of θ, we hve

$\lim_{n\to\infty} P(|t-\theta|<\varepsilon) \to 1$.

To prove $P(|t^2-\theta^2| < \varepsilon) \to 1$. $\left(\begin{array}{c}|x|<\varepsilon \\ -\varepsilon<x<\varepsilon\end{array}\right)$

$\lim_{n\to\infty} P(-\varepsilon < t-\theta < \varepsilon) \to 1$ to avoid θ

$\lim_{n\to\infty} P(-\varepsilon+\theta < t < \theta+\varepsilon) \to 1$

$\lim_{n\to\infty} P(\theta-\varepsilon < t < \theta+\varepsilon) \to 1$ we need t^2.

$\lim_{n\to\infty} P((\theta-\varepsilon)^2 < t^2 < (\theta+\varepsilon)^2) \to 1$

$\lim_{n\to\infty} P(\theta^2 - 2\theta\varepsilon + \varepsilon^2 < t^2 < \theta^2 + 2\theta\varepsilon + \varepsilon^2) \to 1$

we need $t^2-\theta^2$.
so sub θ^2.

$\lim_{n\to\infty} P(\varepsilon^2 - 2\varepsilon\theta < t^2-\theta^2 < \varepsilon^2 + 2\theta\varepsilon) \to 1$ — (1)

Since $\varepsilon^2 - 2\varepsilon\theta < t^2-\theta^2$ $(-2\varepsilon^2)$

$\varepsilon^2 - 2\varepsilon\theta - 2\varepsilon^2 < t^2-\theta^2$

$-\varepsilon^2 - 2\varepsilon\theta < t^2-\theta^2$

$-(\varepsilon^2 + 2\varepsilon\theta) < t^2-\theta^2$ — (2)

Sub (2) in (1)

$\lim_{n\to\infty} P(-(\varepsilon^2+2\varepsilon\theta) < t^2-\theta^2 < \varepsilon^2+2\theta\varepsilon) \to 1$

put $\varepsilon_1 = \varepsilon^2 + 2\varepsilon\theta$.

$\lim_{n\to\infty} P(-\varepsilon_1 < t^2-\theta^2 < \varepsilon_1)$

$\lim_{n\to\infty} P(|t^2-\theta^2| < \varepsilon_1) \to 1$

$\therefore t^2$ is C.(E) for θ^2

(e) Efficiency =

If we have to choose 1 of several unbiased (E) of a given parameter we usually take the 1 whose sample distribution has the smallest variant.

* The (E) with a smaller variance will be concentrated closely to the true value & on the avg., will deviate less from the true value than the (E) with larger variance & they may be considered to be more efficient than the other.

* Let t_1 & t_2 be 2 unbiased (E) of a parameter θ. To choose b/w diff. unbiased (E) 1 would reasonably consider they are variences. (i.e) If $V(t_1)$ is < $V(t_2)$ then t_1 is said to be more efficient than t_2. (i.e) as varience of (E) ↓ses, its efficiency ↑ses.

* $\frac{V(t_1)}{V(t_2)}$ → relative efficiency of t_2 with respect to t_1.

1) If x_1, x_2, x_3 are 3 independent observations from a (P) with mean μ & varience σ^2. If $t_1 = x_1+x_2-x_3$ & $t_2 = 2x_1+3x_2-4x_3$ are 2 estimates of μ. Compare their efficiencies.

A) t_1 & t_2 are unbiased ? → $(t_1 = x_1+x_2-x_3)$

$E(t_1) = E(x_1) + E(x_2) - E(x_3)$

$= \mu + \cancel{\mu - \mu}$

$= \mu$.

$E(t_2) = 2E(x_1) + 3E(x_2) - 4E(x_3)$

$= 2\mu + 3\mu - 4\mu = \mu$

t_1 & t_2 (or) $\therefore \mu = \mu$ → unbiased / consistent v(x) smallest in t_1 / greater in t_2.

$v(t_1) = v(x_1) + v(x_2) + v(x_3)$ | $v(x-y) = v(x)+v(y)$ always true.

$= \sigma^2 + \sigma^2 + \sigma^2 = 3\sigma^2$

$V(t_2) = 2^2 V(x_1) + 3^2 V(x_2) + 4^2 V(x_3)$

$= 2^2\sigma^2 + 3^2\sigma^2 + 4^2\sigma^2$

(σ^2 [illegible] $2^2, 3^2, 4^2$)

$= 4\sigma^2 + 9\sigma^2 + 16\sigma^2 = 29\sigma^2$

$\therefore$ t_1 is more efficient than t_2.

Relative e. of t_2 with r. to t_1,

$$\frac{V(t_1)}{V(t_2)} = \frac{3\sigma^2}{29\sigma^2} = \frac{3}{29}$$

④ Sufficiency =

* An (e) 't' is said to be sufficient if it provides all info contained in the sample in respect of estimating the parameter θ.

* we hve explained the concept of sufficiency of an (e). In many cases a relatively easy criterian for examining an (e) for sufficiency has been developed. This is ⟶ Fisher Neymann factorisation Theorem.

This gives a necessory & sufficient condi- for an (e) 't' to be sufficient for a parameter θ.

⇒ Factorisation Theorem =

let $x_1, x_2 \dots x_n$ be a random sample of size n from a (p) with density () where θ denotes the parameter which may be unknown. Then a statistic $t = t(x_1, x_2, \dots x_n)$ is sufficient if and only if the joint (prob) density () of $x_1, x_2 \dots x_n$ (known as likelyhood of the sample) is capable of being expressed in the form

$$\boxed{L(x_1, x_2 \dots x_n ; \theta) = L_1(t,\theta) \cdot L_2(x_1, x_2 \dots x_n)} \quad ①$$

where the () $L_2(x_1, x_2 \dots x_n)$ is non-ve & does not involve the parameter θ & the () $L_1(t,\theta)$ is non-ve & depending on the parameter θ.

1) S.T the sample mean is sufficient for estimating the parameter λ in poisson distri... ?

A) $f(x,\theta) = f(x,\lambda)$

p. distri ⟶ $\boxed{f(x,\lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}}$, $x = 0, 1, \dots$ $\lambda > 0$

If $x_1, x_2 \dots x_n$ are sample values want to write in the form of ①

$\therefore L(x_1, x_2 \dots x_n, \lambda) = f(x_1,\lambda) \cdot f(x_2,\lambda) \dots f(x_n,\lambda)$

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-\lambda} \cdot e^{-\lambda} \dots e^{-\lambda} \text{ n times } \lambda^{x_1} \cdot \lambda^{x_2} \dots \lambda^{x_n}}{x_1! \cdot x_2! \dots x_n!}$$

$2 \cdot 2 \dots 2 \rightarrow 2^n$

$e^{-\lambda} \cdot e^{-\lambda} \dots e^{-\lambda n} \rightarrow (e^{-\lambda})^n$

$$= \frac{e^{-\lambda n} \cdot \lambda^{x_1 + x_2 + \dots + x_n}}{x_1! \cdot x_2! \dots x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{x_1!\cdots x_n!}$$

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

$n\bar{x} = \frac{1}{y} \cdot \alpha \Sigma x_i$

$= \Sigma x_i$

$$= \frac{e^{-n\lambda}\lambda^{n\lambda}}{x_1!\cdots x_n!} = \lambda^{n\bar{x}} e^{-n\lambda}\left(\frac{1}{x_1!+x_2!\cdots x_n!}\right)$$

$$L(x_1 \cdots x_n ; \lambda) = L_1(\lambda, \bar{x}) \cdot L_2(x_1, x_2 \cdots x_n)$$

$\therefore$ by factorisation (T), the sample mean $\bar{x}$ is a sufficient (e) of λ

2) S.T $\bar{x}$ is a (S) (e) of 'p' when samples of size are taken from binomial (p) with paramters N & p (N gvn).

A) $f(x,p) = N C_x p^x q^{N-x}$ $\quad q = 1-p.$

$= N C_x p^x (1-p)^{N-x} \quad x = 0,1,\ldots N.$

$L(x_1, x_2 \ldots x_n ; p) = f(x,p)$

$= NC_{x_1} p^{x_1}(1-p)^{N-x_1} \cdot NC_{x_2} p^{x_2}(1-p)^{N-x_2} \cdots NC_{x_n} p^{x_n}(1-p)^{N-x_n}$

$= (NC_{x_1})(NC_{x_2}) \cdots (NC_{x_n}) \, p^{\Sigma x_i} (1-p)^{nN - \Sigma x_i}$

$= \underbrace{\text{"} \quad \text{"} \quad \cdots (NC_{x_n})}_{\text{let } L_1} \; \underbrace{p^{n\bar{x}}(1-p)^{nN - n\bar{x}}}_{L_2,}$

$L = L_1 \cdot L_2$ &

$L_1 \rightarrow$ independt of p

$L_2 \rightarrow$ () of $\bar{x}$ & p only.

So $\bar{x}$ is (S) (e) of p

⇒ Completeness =

let $x_1, x_2, \cdots x_n$ denote a random sample from a (p) with density $f(x, \theta)$ & let $t = t(x_1, x_2 \cdots x_n)$ be a statistic then t is said to be complete for θ if & only if the only unbiased (e) of θ, (i.e) a () of t is the statistic that is identically 0 with (prob) 1.

$t = t(x) \quad h(t)$

(i.e) $\boxed{E(h(t)) = 0} \Rightarrow h(t) = 0$ almost surely

(i.e) $P(h(t) = 0) = 1$

eg → let $x_1, x_2 \cdots x_n$ be random sample from bernoulli distri~ gvn by $f(x,p) = p^x(1-p)^{1-x}$ $\quad x = 0,1.$

then $t = \sum_{i=1}^{n} x_i$ is a complete sufficient statistic for p

⇒ Crammer Rao inequality =

let x be a continu~ R.V with pdf $f(x,\theta)$ & the likelyhood () 'L' & let 't' be an unbiased (e) of some () of θ, say $\tau(\theta)$, then

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

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

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

we can write the C-rao inequality as,

$$\boxed{V(t) \geq - \frac{(\tau'(\theta))^2}{E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)}} \quad \text{--- (1)}$$

lower bound

* $\frac{(\tau(\theta))^2}{E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)} \longrightarrow$ C. rao lower bound

known value

Note $\rightarrow$

* If $t = \theta$, then $\tau(\theta) = \theta$ $\therefore$ $\tau'(\theta) = 1$, then eq. (1) becomes,

$$V(t) \geq - \frac{1}{E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)}$$

$\Rightarrow$ **min variance unbiased (E) = (MVUE)**

Let $x_1, x_2 \dots x_n$ be a random sample from (P) with pdf $f(x, \theta)$, Let $t = t(x_1, x_2 \dots x_n)$ be an (E) of θ such that,

a) $E(t) = \theta$ (i.e) t is an unbiased (E)

b) $V(t)$ is < than variance of any other unbiased (E); then t is the min variance unbiased (E) of θ.

$\Rightarrow$ **min variance bounded (E) = (MVBE)**

Let x be a contin- r.v with pdf $f(x, \theta)$ & let t be an unbiased (E) of

same (?) of θ say $\tau(\theta)$, then variance of t is called min v. bounded (MVB) for the (E) of $\tau(\theta)$ & the estimator attaining this bound for every θ is the paramtr space $\longrightarrow$ MVBE

Shiba 10/11/22

1) Let $x_1, x_2 \dots x_n$ be a random sample from normal (P) with unknown mean μ & t non-variance σ^2, find the C. rao lower bond for μ.

A) lower bound $\longrightarrow \frac{(\tau'(\theta))^2}{E\left(\frac{\partial \log L}{\partial \theta}\right)^2}$ --- (1)

normal distri- pdf,

$$f(x, \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty \leq x \leq \infty$$

The likelyhood (L),

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

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

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

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

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

$$= \frac{1}{\sigma^2}\sum(x_i-\mu)$$

$$\left(\frac{\partial \log L}{\partial \mu}\right)^2 = \frac{1}{\sigma^4}\sum(x_i-\mu)^2$$

$$E\left(\frac{\partial \log L}{\partial \mu}\right)^2 = \frac{1}{\sigma^4} E\, n\,(x_i-\mu)^2$$

$$= \frac{n}{\sigma^4} E(x_i - \mu)^2$$

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

$$= \frac{n}{\sigma^4} V(x_i) = \frac{n}{\sigma^4} \cancel{\sigma^2} = \frac{n}{\sigma^2}$$

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

$\tau(\theta) = \theta$.

So $\tau(\mu) = \mu$.

$\tau'(\mu) = 1$

— ①,

$$\Rightarrow \frac{(\tau'(\mu))^2}{E\left(\frac{\partial \log L}{\partial \mu}\right)^2} = \frac{1}{\frac{n}{\sigma^2}} = \frac{\sigma^2}{n}$$

(eg formula)

2) Estimate λ in poisson distri....

$$f(x,\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2 \ldots$$

A) $f(x,\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$

likelyhood (),

$$L = \prod_{i=1}^{n} f(x_i, \lambda) = \frac{e^{-n\lambda} \lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$$

$$\log L = \log (e^{-n\lambda} \lambda^{\Sigma x}) - \Sigma \log x_i!$$

$$= -n\lambda + \sum_{i=1}^{n} x \log \lambda - \Sigma \log_e x_i!$$

$\Sigma x_i = n\bar{x}$

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

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

here $t = \bar{x}$, $t(\theta) = \lambda$

$A(\theta) = \frac{n}{\lambda}$.

$$\frac{\partial \log L}{\partial \theta} = [t - t(\theta)] \cdot A(\theta)$$

$$\boxed{V(t) = \frac{\tau'(\theta)}{A(\theta)}} = \frac{1}{n/\lambda} = \lambda/n$$

hence $\bar{x}$ is MVUE with $V(\theta)$ λ/n.

⇒ Methods of ~~moments~~ Estimation =

a) Methods of moments

b) " of max likelyhood.

a) Methods of moments =

let X be a r.v with pdf $f(x,\theta)$,

let μ_r' be the r^{th} momnt about origin

$$\boxed{\mu_r' = E(x^r)}$$

In general μ_r' will be non () of θ & we write,

$$\boxed{\mu_r' = \mu_r'(\theta)}$$

let $x_1, x_2 \ldots x_n$ be a random sample of size 'n' drawn from the population with pdf $f(x,\theta)$, then r^{th} sample moment will be,

(m → sample)

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

let $m_r' = \mu_r'$ & solve for θ, let $\hat{\theta}$ be the soln of θ, then $\hat{\theta}$ be the estimator of θ obtained by the method of momnts.

(Pro) →

* moment estimators are asymtotically unbiased.
* They are consistnt estimators.
* Under fairly general conditions, the distribu~ moment estimators are

asymtotically normal.

Q
1) Estimate 'p' in the sampling frm binomial (p) $f(x,n,p) = nC_x p^x q^{n-x}$, $x = 0,1,2 \dots n$ by method of momnts?

A) Take a random sample $x_1, x_2 \dots x_n$ frm the binomial (p)

$$\boxed{\begin{array}{l} \mu_r' = E(x) \\ m_r' = \frac{1}{n}\sum x_i \end{array}}$$

$m_1' = \mu_1'$

binomial distrin $E(x) = np$

$m_1' = \frac{1}{n}\sum x_i = \bar{x}$

$np = \bar{x}$

$p = \bar{x}/n$.

Soln $\Rightarrow \hat{p} = \bar{x}/n$

2) Method of max likelyhood =

Suppose that x is a r.v & $x_1, x_2 \dots x_n$ is r.s (sample) of x having the pdf $f(x,\theta)$ also $x_1, x_2 \dots x_n$ are observed sample values then the likelyhood ( ) is

$$L(x_1, x_2 \dots x_n, \theta) = f(x_1, x_2 \dots x_n, \theta)$$
$$= f(x_1,\theta)\, f(x_2,\theta) \dots f(x_n,\theta)$$

$$\boxed{L(x_1, x_2 \dots x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)}$$

The likelyhood ( ) can also be denoted as $L(x,\theta)$ or $L(\theta)$.

The prbm of finding the max likelyhood estimator is the prbm of finding the value of θ that maximises $L(\theta)$.

If there exist a ( ) $t = t(x_1, x_2 \dots x_n)$ of the sample values which maximises L for variations in θ, then t is $\rightarrow$ max likelyhood estimator of θ. (MLE)

Thus t is a soln, if any of $\frac{\partial L}{\partial \theta} = 0$ & $\frac{\partial^2 L}{\partial \theta^2} < 0$.

likelyhood ( ) $\rightarrow L(x,\theta)$

L & $\log L$ have max at the same value of θ. Thus MLE is the soln of the eq.

$$\frac{\partial \log L}{\partial \theta} = 0, \text{ provided } \frac{\partial^2 \log L}{\partial \theta^2} < 0$$

in that case, we must find the value of paramtrs that maximise the likelyhood ( ).

$$\boxed{\begin{array}{l} L(x,\theta) = \prod_{i=1}^{n} f(x_i,\theta) \\ \frac{\partial L}{\partial \theta} = 0 \quad \& \quad \underbrace{\frac{\partial^2 L}{\partial \theta^2} < 0}_{\text{-ve no.}} \end{array}}$$

$\rightarrow$ 2 condi-
$L(x,\theta)$ max

(pro) $\rightarrow$

* MLE are consistnt.
* The distri.n of MLE tends to normality for large ~~dist~~ samples.
* MLE's most efficient.
* MLE are sufficient, if sufficient estimators exist.
* MLE are non-necessarily unbaised.
* MLE have invarience property.

- t MLE of θ then
- $g(t)$ MLE of $g(\theta)$ } invariance (pro).

1) find the max likelyhood (e) of poission paramt λ of poission distri...

A) $f(x,\lambda) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$ $x = 0, 1, \ldots \infty$ $\lambda > 0$.

~~$L(x,\lambda) =$~~

The likelyhood () of random sample $x_1, x_2 \ldots x_n$ of observations from this sample is

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

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{-n\lambda} \cdot \lambda^{x_1} \cdot \lambda^{x_2} \cdots \lambda^{x_n}}{x_1! \cdot x_2! \cdots x_n!}$$

$$L(x,\lambda) = \frac{e^{-n\lambda}\lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$$

2 condi-n

1) $\dfrac{\partial L}{\partial \theta} = 0$, $\dfrac{\partial^2 L}{\partial \theta^2} < 0$

as

2) $\dfrac{\partial \log L}{\partial \theta} = 0$. $\dfrac{\partial^2 \log L}{\partial \theta^2} < 0$. ✓

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

$$= \log e^{-n\lambda} + \log \lambda^{\Sigma x_i} - \sum_{i=1}^{n} \log x_i!$$

$$\log L = -n\lambda + \Sigma x_i \log \lambda - \sum_{i=1}^{n} \log x_i!$$

$$\frac{\partial \log L}{\partial \lambda} = -n + \frac{\Sigma x_i}{\lambda} = 0$$

$\frac{1}{\lambda}\Sigma x_i = n$ $\lambda = ?$

$$\frac{1}{\lambda} = \frac{n}{\Sigma x_i}$$

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

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \frac{\partial}{\partial \lambda}\left(-n + \frac{1}{\lambda}\Sigma x_i\right)$$

$$= -\frac{\Sigma x_i}{\lambda^2} \quad (-ve\ no).$$

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

$\bar{x} = \dfrac{\Sigma x_i}{n}$

$\Sigma x_i = \bar{x} n$

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

$\therefore \dfrac{\partial^2 \log L}{\partial \lambda^2} < 0$

hence $\bar{x}$ is the MLE

⇒ **Least Squares estimators :-**

usually this method is used to fit a curve of the curve $y = f(x_i, a_0, a_1, a_2 \ldots a_n)$ to a where a_i's are unknown paramters set of 'n' sample obsrvations (x_1, y_1) $(x_2, y_2) \ldots (x_n, y_n)$ from a bivariate population. By the method of least squares we minimise the sum of squares of the error b/w the obsrved & estimated values gvn by,

$$S = \sum_{i=1}^{n}\left(y_i - f(x_i, a_0, a_1, \ldots a_n)\right)^2$$

Subject to variations in $(a_0, a_1, \dots a_n)$ for 's' is to be min we hve

$\boxed{\frac{\partial S}{\partial a_i} = 0}$ for $i = 0, 1, 2 \dots n$

This are $\longrightarrow$ Normal eqs.

Solving this eqs, we gt the least square (e) $a_0, a_1, \dots a_n$.

⇒ Method of min Varience :-

Let $x_1, x_2 \dots x_n$ be a r.sample drawn from a (p) with pdf $f(x, \theta)$, then the likelyhood (L) is,

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

The prbm is to find a statistic $t = t(x_1, x_2 \dots x_n)$ such that

$$E(t) = \int_{-\infty}^{\infty} t \, L \, dx = \gamma(\theta)$$

(gamma)

$$\Rightarrow E(t) = \int_{-\infty}^{\infty} (t - \gamma(\theta)) \, L \, dx = 0 \quad \text{--- (1)}$$

$$\Rightarrow V(t) = \int_{-\infty}^{\infty} (t - E(t))^2 \, L \, dx$$

$$= \int_{-\infty}^{\infty} (t - \gamma(\theta))^2 \, L \, dx \quad \text{--- (2)} \text{ is min}$$

In other words, we hve to minimise --- (2) subject to the condit --- (1)

here $\int_{-\infty}^{\infty} dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots dx_1 \, dx_2 \cdots dx_n$