* Estimation :

Estimate — single best guess of the value of parameter.

Parameter — Population constant (μ, σ etc).

Estimator : — refers to a statistic (constant) that is used to generate an estimate once data is collected. Estimator can also be thought as a 'rule' that creates an estimate. eg. sample variance is estimator of pop variance.

(Mean of sampling distribution of means = μ = pop. mean.)

~~Types of~~

- Estimation is concerned with the methods by which population characteristics are estimated from sample information.

(True value of parameter is unknown which can be correctly obtained by exhaustive study of population which is expensive & might be infeasible) -

statistical estimation procedures provide us with the means of obtaining estimates of population parameters with desired degrees of precision.

- Two types of estimates
  - Point estimate — single number
  - Interval estimate — range of pop. parameter.

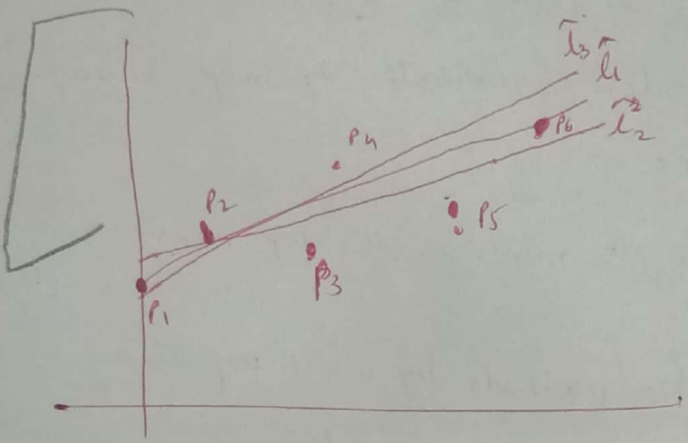
Point estimate : — A single no. which is used as an estimate of of the unknown population parameter.

The procedure in point estimation is to select a random sample from n observations. from a population $f(x, \theta)$ & then use some preconceived method to arrive from these observations at a no. say $\hat{\theta}$ which is an estimator of θ.

Properties of good estimator:

(i) Unbiasedness:-

An estimator $\hat{\theta}$ for an unknown model parameter θ is said to be unbiased estimate of θ if $E(\hat{\theta}) = \theta$.



$y = \alpha + \beta x + e_i$

$\hat{\beta}_1\ \hat{\beta}_2\ \hat{\beta}_3$ - slopes of $\hat{l}_1\ \hat{l}_2\ \hat{l}_3$.

$\hat{\beta}_1$ - slope of line $\hat{l}_1$ joining pts P_1 & P_6.

$\hat{\beta}_2$ - slope of line $\hat{l}_2$ joining midpts of $P_1 P_2$ & midpt $P_5 P_6$.

$\hat{\beta}_3$ - slope of line joining centre of gravity of $P_1 P_2 P_3$ to centre of gravity $P_4 P_5 P_6$ (c.G. - $\frac{1}{3}(P_1+P_2+P_3)$ (centroid)).

Model: $Y_i = \alpha + \beta x_i + e_i$

$\hat{\beta}_1 = \frac{y_6 - y_1}{x_6 - x_1}$ or $\hat{\beta}_1 = \frac{Y_6 - Y_1}{x_6 - x_1}$

$Y_6 = \alpha + \hat{\beta}_1 x_6$ $\quad \hat{Y}_6 = \beta_1 x_6$

$Y_5 = \alpha + \hat{\beta}_1 x_5$ $\quad Y_5 = \beta_1 x_5$

Now $E(Y_i) = E(\alpha + \beta x_i + e_i)$
$= \alpha + \beta x_i$

$E(e_i) = 0$

$E(\hat{\beta}_1) = \frac{E(Y_6 - Y_1)}{x_6 - x_1} = \frac{1}{x_6 - x_1}(\alpha + \beta x_6 - \alpha - \beta x_1)$
$= \frac{\beta(x_6 - x_1)}{x_6 - x_1} = \beta$

$\Rightarrow E(\hat{\beta}_1) = \beta$

$y = \alpha + \beta x$

for line $\hat{l}_1$ estimate is

$y_i = \alpha + \hat{\beta}_1 x_i$

$E(\hat{\beta}_1) = E\left(\frac{Y_6 - Y_1}{x_6 - x_1}\right)$

$= E\left(\frac{\alpha + \hat{\beta}_1 x_6 - \alpha - \hat{\beta}_1 x_1}{x_6 - x_1}\right)$

$= \beta$.

not required.

(ii) ~~Consistency~~:- Unbiasedness is a statement about the expected value of the sampling distribution of the estimator.

Ex: Unbiasedness

Consider a sample of n independent draws from a normal distribution having unknown μ & variance σ^2.

As an estimator of the mean μ, we use the sample mean,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i . \qquad \text{To prove } E(\bar{X}_n) = \mu .$$

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

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n}{n}\mu \qquad \because E(X_i)\ \forall i \text{ in normal dist} = \underline{\underline{\mu}}.\ \text{mean}$$

$$E(\bar{X}_n) = \mu . \qquad \Rightarrow \bar{X}_n \text{ is unbiased.}$$

- Variance of estimator $\bar{X}_n = \sigma^2/n$.

$$\because Var(\bar{X}_n) = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] \qquad (\because V(aX) = a^2 V(X))$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\left[Var(X_i)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

$\therefore$ Variance of estimator $\to 0$ as sample size $n \to \infty$.

(2) To prove variance as unbiased estimator for a normal dist^n with known mean 'μ':

- For variance estimator of variance $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

true variance $= \sigma^2$

Estimate $\hat{\sigma}_n^2$ is unbiased if $E[\hat{\sigma}_n^2] = \sigma^2$.

$$\rightarrow E[\hat{\sigma}_n^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right].$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X_i - \mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} Var(X_i) \qquad (\text{by def } V(X) = E(X - \bar{X})^2 \text{ or } \underline{E(X-\mu)^2}$$

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

$$\therefore E[\hat{\sigma}_n^2] = \sigma^2 \Rightarrow \hat{\sigma}_n^2 \text{ is unbiased estimator of } \sigma^2.$$

② **Consistency:-** If an estimator $\hat{\theta}$ approaches the parameter θ closer & closer as the sample size 'n' increases, $\hat{\theta}$ is said to be a consistent estimator of θ.

Estimator $\hat{\theta}$ is said to be a consistent estimator of θ if, as n approaches infinity to probability approaches 1.

In case of large samples consistency is a desirable property for an estimate to possess.

– Consistency of estimator means that as the sample size gets large the estimate gets closer & closer to the true value of the parameter.

Sample mean & Sample variance are consistent estimates.

③ **Efficiency:-** The concept of efficiency refers to the sampling variability of an estimator. If two competing estimators are both unbiased, the one with smaller variance (for a given sample size) is said to be relatively more efficient.

" Estimator $\hat{\theta}_1$ is said to be more efficient than another estimator $\hat{\theta}_2$ for θ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$".

The smaller the variance ~~of the first is less than the variance of the second~~. of the estimator, the more concentrated is the distribution of the estimate around the parameter being estimated.

* If the population is symmetrically distributed, then both the sample mean & median are consistent & unbiased estimators of μ. But mean is more efficient than the median.

The biasedness can be corrected by modifying estimator '$\hat{\theta}_1$' as

$$\hat{\theta}_2 = \frac{n+1}{n} \cdot \max(X_1, X_2, \cdots X_n)$$

$$E(\hat{\theta}_2) = \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta.$$

* (3) Can a sample ~~proportion~~ mean $\frac{\Sigma X_i}{n}$ be considered as an unbiased estimator of λ, where ~~the~~ X is a r.v. for poisson distⁿ with parameter λ. $X = \Sigma X_i$

→ Let $\hat{\lambda} = \frac{\Sigma X_i}{n}$

$$E(\hat{\lambda}) = E\left[\frac{\Sigma X_i}{n}\right] = \frac{1}{n} E(\Sigma X_i)$$
$$= \frac{1}{n} \Sigma E(X_i)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \lambda$$
$$= \frac{n\lambda}{n} = \lambda.$$

* Proof that median is unbiased estimator for normal distⁿ with parameter μ & σ^2.

Let $X_1, X_2, \cdots X_n$ be the sample (To prove $E[\text{median}(X_1, X_2 \ldots X_n)] = \mu$)

let $Y_i = X_i - \mu$ for $i = 1, 2, \cdots n$.

Let $m = E(\text{median}) = E[\text{median}(Y_1, Y_2, \cdots Y_n)]$.

∵ Normal distⁿ is symmetric about $Y = 0$;

$-m = E(-\text{median}) = E(\text{median}) = m$.

∴ $-m = m \Rightarrow m = 0. \Rightarrow E(\text{median}) = 0.$

$$\Rightarrow E[\text{median}(X_1, X_2, \ldots X_n)] = E[\text{median}\{Y_1 + \mu, Y_2 + \mu, \cdots Y_n + \mu\}]$$
$$= E[\mu + \text{median}(Y_1, Y_2, \cdots Y_n)]$$
$$= E(\mu) = \mu.$$

$(Z_i = -Y_i \Rightarrow m = E(\text{median}(Y_1, Y_2, \cdots Y_n)) = E(\text{median}(Z_1, \cdots Z_n))$
$$= E(-\text{median}(Y_1 Y_2 \cdots Y_n))$$
$$= -m \Rightarrow m = -m$$
$$\Rightarrow m = 0$$

Scanned by CamScanner

Ex:

① Can a sample proportion $\frac{X}{n}$ be considered as an unbiased estimator of p, where the random variable X is the no. of successes has binomial distⁿ with parameters n & p. Given $\hat{p} = \frac{X}{n}$

→ $E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n}(np) = p.$

② Suppose that X, the reaction time to a certain stimulus has a uniform distribution on the interval from 0 to an unknown upper limit θ

$\therefore f(x) = \frac{1}{\theta} \quad 0 \le x \le \theta.$

→ It is desired to estimate θ on the basis of a random sample $X_1, X_2, \ldots X_n$ of reaction times.

$\because$ θ is the largest possible time in the entire population of reaction times, consider as the first estimator the largest sample reaction time a say $\hat{\theta}_1 = \max(X_1, X_2, \ldots X_n)$

for ex. if $n=5$, $x_1 = 4.2$, $x_2 = 1.7$, $x_3 = 2.4$, $x_4 = 3.9$, $x_5 = 1.3$

The point estimate of θ is

$$\hat{\theta}_1 = \max(4.2, 1.7, 2.4, 3.9, 1.3) = 4.2.$$

Now unbiasedness implies that some samples will yield estimates that exceed θ & other samples will yield estimates smaller than θ otherwise θ could not be the centre of $\hat{\theta}_1$'s distribution.

However, our proposed estimator will never overestimate θ & will ~~always~~ underestimate θ unless the largest sample value equals θ.

⇒ $\hat{\theta}_1$ is a ~~bised~~ biased estimator.

Infact it can be shown that $E(\hat{\theta}_1) < \theta$.

$$E(\hat{\theta}_1) = \frac{n}{n+1} \cdot \theta < \theta \qquad \because \frac{n}{n+1} < 1.$$

But this estimate is consistent as when $n \to \infty$ $\hat{\theta}_1 \to \theta$.

$E(\hat{\theta}_1) = \theta.$

(A) Sufficiency :-

An estimator is said to be sufficient if it conveys as much information as is possible about the parameter which is contained in the sample.

The significance of sufficiency lies in the fact that if a sufficient estimator exists, it is absolutely unnecessary to consider any other estimator.

A sufficient estimator ensures that all information a sample can furnish w.r. to the estimation of a parameter being utilised.

Efficiency : An efficient estimator is the 'best possible' or optimal estimator of a parameter of interest

An estimator is said to be efficient if in the class of unbiased estimators it has min. variance

Eg: Consider a sample from normal distⁿ.
For normal distⁿ sample mean $\bar{x}$ & sample median are unbiased.

$\therefore E(\bar{x}) = \mu$ & $E(\text{median}) = \mu$.

Now $V(\bar{x}) = \frac{\sigma^2}{n}$, $V(\text{median}) = \frac{\pi}{2} \cdot \frac{\sigma^2}{n}$.

$\Rightarrow \bar{x}$ is more efficient $\because V(\bar{x}) < V(\text{median})$.

Scanned by CamScanner

Sufficiency :- Estimator is sufficient if it uses all the sample information.

For normal distⁿ → mean & median are unbiased estimators.

However median uses only rank so it is not sufficient while sample mean considers each member of the sample as well as its size so it is sufficient statistic.

Methods of estimation —

① -

Methods of estimation –

1. Bayes estimators, ② least square estimator ③ Method of moments estimator

4. Maximum likelihood estimators. 5. Minimum mean squared error etc.

① Method of Least Squares –

Suppose $x_1, x_2, \dots x_n$ is a random sample of the population whose parameter θ is the mean of the population which is unknown.

~~Then each x_i should~~ .

Then the reasonable estimate $\hat{\theta}$ of θ can be ~~considered as~~ found by considering the sum of squares of the differences as small as possible.

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

where $(x_i - \theta)$ is the error term in fitting the regression line.

The concept of minimizing the sum of squared differences betn observed data & expected data is known as the principle of least squares.

Ex:- Suppose observations $x_1 = 3, x_2 = 4, x_3 = 8$ are collected in a random sample of size 3 from a population with unknown mean θ.

Then $S = \sum_{i=1}^{3} (x_i - \theta)^2 = (3-\theta)^2 + (4-\theta)^2 + (8-\theta)^2$

$= 89 - 30\theta + 3\theta^2$.

S is a fn of θ. To identify the value of θ for which the sum is minimized we consider

θ	0	1	2	3	4	5	6	7	8 ...
S	89	62	41	26	17	14	17	26	41 ...

⇒ for low values of $\theta = 0, 1, 2$ ⇒ S is high & it attains min at $\theta = 5$.

where $\theta = 5$ is also a mean of x_1, x_2, x_3

⇒ Sample mean $\bar{x} = 5$ is required estimate of the unknown population mean.

In General, for a random sample of size n taken from a ~~populat~~ population with unknown mean θ, the expression of sum of squares is given as

$$S = \sum_{i=1}^{n} (x_i - \theta)^2 = \sum_{i=1}^{n} (x_i^2 - 2\theta x_i + \theta^2)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\theta \sum_{i=1}^{n} x_i + n\theta^2.$$

S is a quadratic f^n of θ.

To minimize the sum, differentiate S w.r.to θ, we get

$$\frac{dS}{d\theta} = 0 - 2\sum x_i + 2n\theta$$

$$\& \ \frac{dS}{d\theta} = 0 \implies \frac{\sum_{i=1}^{n} x_i}{n} = \theta \implies \theta = \bar{X}$$

$\implies$ Sample mean is a least sq. estimate of pop mean θ.

Metho of moments :

The method of moments involves equating sample moments with theoretical moments.

① $E(X^k)$ = k^{th} theoretical moment of distribution about origin

② $E[(X-\mu)^k]$ → k^{th} theoretical moment of distribution about mean.

③ $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ is k^{th} sample moment

④ $M_k^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ - is k^{th} sample moment about sample mean.

The basic concept of this method is

① Equate 1^{st} sample moment about origin to 1^{st} theoretical moment i.e $M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ to $E(X)$.

② Equate 2^{nd} sample moment about origin to 2^{nd} theoretical moment $E(X^2)$ & so on.

③ Continue equating sample moments about origin ~~too~~ with the corresponding theoretical moments $E(X^k)$ until we get the ~~required~~ no. of equations to be equal to the no. of parameters.

④ Solve these eq^{n}s for parameters.

Resulting values are called method of moments estimators.

Ex: 1

Let $X_1, X_2, \ldots X_n$ be Binomial variables with parameter p.

what is the method of moments estimator of p?

(1) $E(X) = p$. $\rightarrow$ Theoretical moment

As there is only one parameter we need only one eqⁿ.

Equating 1st Theoretical momement about origin with the corresponding sample moment we get

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

$$\Rightarrow \hat{p}_{MM} = \frac{1}{n}\sum_{i=1}^{n} x_i$$ is the required moments.

(2) ||^{ly} for Poisson distⁿ $\hat{\lambda}_{MM} = \bar{x}$.

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

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

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

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

not required.

(3) Let $X_1, X_2, \ldots X_n$ be normal random variables with mean μ & variance σ^2. Calculate the method of moments estimators of the mean μ & variance σ^2.

$\rightarrow$ The 1st & 2nd theoretical moments about origin are

$$E(X) = \mu \quad \& \quad E(X^2) = \sigma^2 + \mu^2. \quad (\sigma^2 = E(X^2) - (E(X))^2$$

. Here we have 2 parameters. so we need 2 equations.
· Equating 1st theoretical moment to 1st sample moment we get

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

2nd $\rightarrow$ $E(x^2) = \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 . \rightarrow (2)$

from eqn (1);

$$\hat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} .$$

Substituting the sample mean for μ in eqn (2) & solving for σ^2

we get method of moments estimator for variance σ^2 as —

$$\hat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 .$$

$$\boxed{\therefore \hat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\left(\Sigma x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)$$

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

$$\frac{\Sigma x_i}{n} \cdot \frac{\Sigma x_i}{n}$$

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

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

$$\frac{1}{n}\Sigma x_i^2 - 2\bar{x}^2 + \bar{x}^2$$

$$\frac{1}{n}\Sigma x_i^2 - 2\frac{\Sigma x_i}{n}\cancel{\bar{x}} + \frac{\Sigma x_i}{n}$$

$$\frac{1}{n}\Sigma (x_i - \bar{x})^2 = \frac{1}{n}\left[\Sigma\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\right.$$

$$\frac{\Sigma x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2$$

Σr

Scanned by CamScanner

Method of maximum likelihood estimation.

Ex: Descripancy using method of moments —

~~Sq~~ 3 observations were collected on a continuous uniform r.v. $X \sim U(0, \theta)$ where parameter θ is unknown. The aim of the sample is to estimate θ. The data recorded were $x_1 = 3.2$, $x_2 = 2.9$, $x_3 = 13.1$

Sample mean $= \bar{x} = 6.4$.

mean of uniform distⁿ $= \frac{b-a}{2} = \frac{\theta}{2}$.

$\therefore$ By method of moments, $\bar{x} = \frac{1}{2}\hat{\theta}$.

$\Rightarrow \hat{\theta} = 2\bar{x} = 12.8$.

$\Rightarrow$ which means that the prob. model defined must take values only over the range $[0, 12.8]$ & it will not permit $x_3 = 13.1$ yet that was the value obtained in the sample.

* Consider an example of censored data —

An area of soil was divided into 240 regions of equal area 'called quadrants'. & in each quadrant the no. of colonies of bacteria found was counted. The data are given in the table below:

Count :	0	1	2	3	4	5	≥ 6
Freq :	11	37	64	55	37	24	12

Here precise record about the sample is not kept

$\therefore$ method of moments can not be applied since sample mean can not be calculated.

In such cases, the method of maximum likelihood is adopted.

Defn: If several independent observations $x_1, x_2, \dots x_n$ are collected on the discrete r.v. X with pmf

$P_X(x) = p(x; \theta)$ — $x = 0, 1, 2, \dots$ where θ is the parameter to be estimated.

Then the product

$$P(x_1, x_2, \dots x_n; \theta) = P(x_1; \theta) \times P(x_2; \theta) \times \dots \times P(x_n; \theta)$$

is known as the likelihood of θ for the random sample $x_1, x_2, \dots x_n$.

The value $\hat{\theta}$ of θ at which the likelihood is maximised is known as maximum likelihood estimator of θ.

whereas for continuous r.v. X; the ~~maximum~~ likelihood of θ for a random sample $x_1, x_2, \dots x_n$ is given as the product

$$f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta).$$

In some cases instead of product log likelihood is taken then we have

$$\text{log likelihood } \theta = l(\theta) = \log [f(x_1; \theta) \cdot f(x_2; \theta) \cdots f(x_n; \theta)]$$

$$= \sum_{i=1}^{n} \log f(x_i; \theta).$$

Scanned by CamScanner

Estimation (unit 2)

~~Ichthen data~~

Maximum Likelihood Estimation:

The max. likelihood estimate (mle) of θ is that value of θ that maximises $lik(\theta)$; i.e. it is the value that makes the observed data the "most probable." If X_i's are iid then $lik(\theta) = \prod_{i=1}^{n} f(x_i/\theta)$

log likelihood $= l(\theta) = \sum_{i=1}^{n} \log(f(x_i/\theta))$

Poisson Example:- (Estimating Poisson parameter)

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

- For $X_1, X_2, \dots X_n$ iid (independent & identically distributed r.vs) Poisson r.vs will have a joint freq. f^n that is a product of the marginal freq. functions. $l(\theta) = \sum_i \log f(x_i;\theta) =$

log likelihood of poisson distribution is

$$l(\lambda) = \sum_{i=1}^{n} (x_i \log\lambda - \lambda - \log x_i!)$$

replace λ by θ.

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

- To maximize this f^n, we find derivative (2nd derivative test)

$$l'(\lambda) = \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i - n \quad \& \quad l'(\lambda) = 0$$

$$\Rightarrow \lambda = \frac{\sum x_i}{n} = \bar{x}.$$

$\Rightarrow$ estimate is $\underline{\underline{\hat{\lambda} = \bar{x}}}$ $\left(\theta = \bar{x} \Rightarrow \hat{\theta} = \hat{\lambda} = \bar{x}.\right)$

Scanned by CamScanner

Estimating the exponential parameter.

For a random sample $x_1, x_2, \cdots x_n$ from an exponential dist^n with unknown parameter θ, the corresponding pdf is

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

$\therefore$ likelihood of θ of the sample is

$$l(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdots f(x_n;\theta).$$

$$= \theta e^{-\theta x_1} \cdot \theta e^{-\theta x_2} \cdots \theta e^{-\theta x_n}$$

$$= \theta^n e^{-\theta(x_1+x_2 \cdots + x_n)}. \quad = \theta^n e^{-\theta \Sigma x_i}$$

Log likelihood
$l(\theta) = n \log\theta - \theta \Sigma x_i$
$\therefore \frac{dl(\theta)}{d\theta} = \frac{n}{\theta} - \Sigma x_i \Rightarrow \hat{\theta} = \frac{1}{\bar{x}}$

$$\frac{d(l(\theta))}{d\theta} = n\theta^{n-1} e^{-\theta \Sigma x_i} - \theta^n \Sigma x_i e^{-\theta \Sigma x_i}$$

$$\therefore \frac{d\,l(\theta)}{d\theta} = 0 \Rightarrow \theta^{n-1} e^{-\theta(\Sigma x_i)} [n - \theta \Sigma x_i] = 0$$

$\neq 0$.

$\Rightarrow n - \theta \Sigma x_i = 0$

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

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

mle $= \hat{\theta}$

$\therefore$ mle $\theta = \hat{\theta} = \frac{1}{\text{mean}}$

Example 8.9

For the following random samples, find the maximum likelihood estimate of θ:

1. $X_i \sim Binomial(3, \theta)$, and we have observed $(x_1, x_2, x_3, x_4) = (1, 3, 2, 2)$.
2. $X_i \sim Exponential(\theta)$ and we have observed $(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$.

- Solution
  - 1. In Example 8.8., we found the likelihood function as

$$L(1, 3, 2, 2; \theta) = 27 \quad \theta^8 (1-\theta)^4.$$

To find the value of θ that maximizes the likelihood function, we can take the derivative and set it to zero. We have

$$\frac{dL(1, 3, 2, 2; \theta)}{d\theta} = 27 \big[\quad 8\theta^7 (1-\theta)^4 - 4\theta^8 (1-\theta)^3 \big].$$

Thus, we obtain

$$\hat{\theta}_{ML} = \frac{2}{3}.$$

2. In Example 8.8., we found the likelihood function as

$$L(1.23, 3.32, 1.98, 2.12; \theta) = \theta^4 e^{-8.65\theta}.$$

Here, it is easier to work with the log likelihood function, $\ln L(1.23, 3.32, 1.98, 2.12; \theta)$. Specifically,

$$\ln L(1.23, 3.32, 1.98, 2.12; \theta) = 4 \ln \theta - 8.65\theta.$$

By differentiating, we obtain

$$\frac{4}{\theta} - 8.65 = 0,$$

which results in

$$\hat{\theta}_{ML} = 0.46$$

It is worth noting that technically, we need to look at the second derivatives and endpoints to make sure that maximum likelihood estimators $\Theta_1, \Theta_2, \cdots, \Theta_k$.

Example 8.11

Suppose that we have observed the random sample $X_1, X_2, X_3, \ldots, X_n$, where $X_i \sim N(\theta_1, \theta_2)$, so

$$f_{X_i}(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}}.$$

Find the maximum likelihood estimators for θ_1 and θ_2.

- Solution
  - The likelihood function is given by

$$L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2) = \frac{1}{(2\pi)^{\frac{n}{2}} {\theta_2}^{\frac{n}{2}}} \exp\left(-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2\right).$$

Here again, it is easier to work with the log likelihood function

$$\ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\theta_2 - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2.$$

We take the derivatives with respect to θ_1 and θ_2 and set them to zero:

$$\frac{\partial}{\partial\theta_1}\ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2) = \frac{1}{\theta_2}\sum_{i=1}^{n}(x_i-\theta_1) = 0$$

$$\frac{\partial}{\partial\theta_2}\ln L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(x_i-\theta_1)^2 = 0.$$

By solving the above equations, we obtain the following maximum likelihood estimates for θ_1 and θ_2:

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\theta_1)^2.$$

We can write the MLE of θ_1 and θ_2 as random variables $\hat{\Theta}_1$ and $\hat{\Theta}_2$:

$$\hat{\Theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\hat{\Theta}_2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \Theta_1)^2.$$

← previous
next →

let $X_1, X_2, \cdots X_n$ be a random sample from a uniform $U(0,\theta)$ distribution, where θ is unknown parameter. Find MLE of θ based on this sample.

→ If $X_i \sim U(0,\theta)$ then,

$$f_X(x) = \frac{1}{\theta} \qquad 0 < x \leq \theta$$
$$= 0 \qquad \text{otherwise}.$$

The likelihood f^n is given by

$$L(x_1, x_2, \cdots x_n; \theta) = f_{X_1, X_2, \cdots X_n}(x_1, x_2, x_3 \cdots x_n; \theta)$$
$$= f_{X_1}(x_1;\theta) \cdot f_{X_2}(x_2;\theta) \cdot \cdots \cdot f_{X_n}(x_n;\theta)$$
$$= \frac{1}{\theta^n} \qquad 0 \leq x_1, x_2, \cdots x_n < \theta$$
$$= 0 \qquad \text{o.w.}$$

Now $\frac{1}{\theta^n}$ is decreasing f^n of θ. Thus to maximize it, we need to choose the smallest possible value for θ.

But for $i = 1, 2, \dots n$; $x_i \in (0, \theta)$.

$\Rightarrow$ Smallest possible value of θ is

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

$\therefore$ MLE of θ is

$$\hat{\theta}_{MLE} = \max(x_1, x_2, \dots x_n).$$

In this case $\hat{\theta}_{MLE}$ can not be obtained by setting the derivative of the likelihood f^n to zero.

The maximum is achieved at an endpoint of the acceptable interval.

Scanned by CamScanner