

Rao - Blackwell Theorem (RBT)

and MUVE sufficient statistics play an important role in finding good estimators for parameters

If $\hat{\theta}$ is an unbiased estimator for θ & if U is a statistic that is sufficient for θ .

Then, there is a function of U that is also an unbiased estimator for θ and has no larger variance than $\hat{\theta}$.

If we seek unbiased estimators with small variances we can restrict our research to estimators that are functions of sufficient statistics

RBT states that,

Let $\hat{\theta}$ be a unbiased estimator of θ such that $V(\hat{\theta}) < \infty$
If U is a sufficient statistic for parameter θ . Define

$$\hat{\theta}^* = E(\hat{\theta} | U) \text{ then for all } \theta \\ E(\hat{\theta}^*) = \theta \text{ and } V(\hat{\theta}^*) \leq V(\hat{\theta})$$

RBT implies that an unbiased estimator for $\hat{\theta}$ with small Variance is or can be made to be a function of sufficient statistic

If we start with an unbiased estimator for parameter θ and the sufficient statistic is obtained through factorization criterion, application of RBT leads to a MUVE (minimum variance unbiased estimator) for the parameter

Unbiased
Efficient
sufficient } these statistics passes
property known as
Completeness

If U is a sufficient statistic that best summarizes the data and some function of U such ~~that~~ as $h(U)$ can be found such that

$$E[h(U)] = \theta$$

It follows that $\rightarrow h(U)$ is MUVE for θ

Ex: Consider a bernoulli rv. with unknown p . Use factorization theorem to find a sufficient statistic that best summarizes data. Give a MUUE for p

$$P(Y_i = y_i) = p^{y_i} (1-p)^{1-y_i}, y_i = 0, 1.$$

$$\begin{aligned} L(p) &= L(y_1, y_2, \dots, y_n | p) = P(y_1, y_2, \dots, y_n | p) \\ &= p^{y_1} (1-p)^{1-y_1} \cdot p^{y_2} (1-p)^{1-y_2} \dots p^{y_n} (1-p)^{1-y_n} \\ &= \frac{p^{\sum y_i} (1-p)^{n - \sum y_i}}{g(\sum y_i, p)} \cdot \underbrace{1}_{h(y_1, y_2, \dots, y_n)} \end{aligned}$$

According to factorization theorem

$$U = \sum_{i=1}^n y_i \text{ is a sufficient statistic for } p$$

$$E(U) = np \text{ or } E(U/N) = p$$

$U/N = \bar{y}$ is an unbiased estimator for p

Since this estimator is a function of sufficient statistic

$\sum y_i \rightarrow$ estimator $\hat{p} = \bar{y}$ is a MUUE for p

Sufficient + unbiased $\Rightarrow$ MUUE

Methods for finding Point Estimators

- $\rightarrow$ method of moments (mm)
- $\rightarrow$ method of maximum likelihood (ML)
- $\rightarrow$ least squares method (LS)
- $\rightarrow$ Bayesian Estimation

Estimator & Estimate

Let Y be a rv. with pdf $f(y | \theta)$ where θ is a parameter we want to estimate.

Let $y_1, y_2, \dots, y_n$ be a random sample from this population.

Then an estimator of θ is a function or rule of the form

$$\hat{\theta} = \hat{\theta}(y_1, y_2, \dots, y_n)$$

which signifies that estimator $\hat{\theta}$ is a function of $y_1, y_2, \dots, y_n$.

Estimator $\hat{\theta}$ is a random variable since its a function of random variables

When values of rv's are inserted into equation $\rightarrow$ an estimate is produced

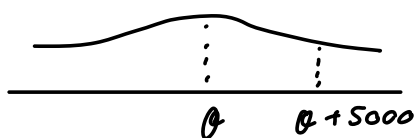
Estimate $\rightarrow$ is value of random variable or estimator $\hat{\theta}$

$\downarrow$
It is not random

An estimator is a function of sample observations that does not depend on any unknown parameters

Ex: Suppose we assume that expenditures by individuals with an annual income of 20,000 \$ are approximately normally distributed with an unknown mean θ but known variance $\theta^2 = 500^2$ that is X is a random variable denoting expenditure from a \$20,000 income by a person in this population

$$Y \sim N(\theta, \sigma^2 = 500^2)$$



Let $y_1, y_2, \dots, y_n$ be a random sample of observations on expenditures from this population. Thus, we are asserting that

$$y_i = \theta + e_i, \quad i = 1, \dots, n$$

where random var. $e_i = y_i - \theta$ has distribution

$$e_i \sim N(0, 500^2)$$

represents difference between y_i & its mean. e_i represents all factors that influence level of expenditure other than income

$y_i = \theta + e_i \Rightarrow$ is a linear statistical model that represents our maintained hypothesis about underlying sampling process or how sample observations on expenditure are obtained

Sample observations are assumed to be experimental outcomes equal mean plus some random disturbance, e_i well defined distr.

A common estimator of θ is an arithmetic mean

$$\hat{\theta} = \hat{\theta}(y_1, \dots, y_n) = \sum y_i / n$$

$\hat{\theta}$ is a random variable since rule says no matter what values of y_i are sum them & divide by sample size

If $n=3$, $y_1=9000$, $y_2=12000$, $y_3=18000$
parameter estimate is

$$\hat{\theta} = \sum_{i=1}^3 y_i / n = 13000$$

Method of Moments (MM)

r^{th} moment of a rv Y about origin is

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

If pdf of Y is $f(y|\theta)$ where $\theta' = (\theta_1, \dots, \theta_k)$ is a vector of unknown parameters, then in general μ_r' will be a known function of θ ,

$$\mu_r' = \mu_r'(\theta)$$

Idea of method of moment is to use random sample of data $y_1, y_2, \dots, y_n$ to compute sample moments

$$\hat{\mu}_r' = \sum_{i=1}^n y_i^r / n, \quad r=1, 2, \dots, k$$

then equate sample & true moments

$$\hat{\mu}_r' = \mu_r'(\theta)$$

and solve resulting system of k equations for unknown parameters if possible

Resulting estimator, $\hat{\theta}$ is method of moments estimator.

Ex: Let $Y_1, \dots, Y_n$ be a random sample for a $N(0, \sigma^2)$ population. Recall that

$$E(Y) = 0 = \mu_1, \quad \sigma^2 = E(Y^2) - [E(Y)]^2 \\ = \mu_2' - (\mu_1')^2$$

$$\hat{\mu}_1' = \sum_{i=1}^n y_i / n = \bar{y} = \hat{\theta}$$

$$\hat{\mu}_2' = \sum_{i=1}^n y_i^2 / n = \hat{\sigma}^2 + \hat{\theta}^2$$

$$\text{So, } \hat{\sigma}^2 = \sum y_i^2 / n - \bar{y}^2 = \boxed{1/n \sum (Y_i - \bar{y})^2}$$

There are difficulties of this method

- 1- These estimators need to be unique
- 2- Method can be difficult to apply in more complicated problems
- 3- It also depends on random variables in question having moments, which is not always the case.

Method of Maximum Likelihood (ML) or (MLE)

Consider Y that is a Bernoulli random variable with parameter $p \rightarrow 1/4$
 $p \rightarrow 3/4$

Suppose we have a random sample of size $n=3$ with values $y_1=1, y_2=1, y_3=0$

Question is how to use these data to estimate unknown parameter p

Method of ML will choose that value of unknown parameter p that maximizes probability (likelihood) of randomly drawing a sample that was actually obtained

Since $f(y|p) = p^y (1-p)^{1-y}$ $y=0,1$

we can calculate probability of our random sample from joint pdf of Y_1, Y_2, Y_3

$$f(y_1=1, y_2=1, y_3=0) = f(1,1,0) = \prod_{i=1}^3 p^{y_i} (1-p)^{1-y_i} \\ = p \cdot p \cdot (1-p)$$

Is a function of p given sample observations. Its called likelihood function and written as $l(p|y)$

Likelihood function is mathematically identical to joint pdf of random sample

But it is interpreted as a function of unknown parameters instead of values of random variables

In making inferences or decisions about p after sample values are observed, all relevant sample information is contained in likelihood function

$$l(114|y) = 0.0468 \quad l(314|y) = 0.1406$$

Thus, likelihood of obtaining sample values we actually have is maximized by choosing $\hat{p} = 314$ rather than $\hat{p} = 114$

$$\hat{p} = 314 \rightarrow \text{MLE of } p$$

It is the value of p most likely to have generated our sample however, it will rarely be case that set of parameters contains only two elements

p can only take any value in parameter space

$$\Omega = \{p | 0 \leq p < 1\}$$

one way is to use numerical technique to maximize

$$\max l(p|y) = p^2(1-p)$$

Second way is to find value of p that maximizes

$\max l(p|y)$ by using calculus techniques

A local max of continuous function occurs $\rightarrow$ its slope = 0
2nd derivative negative

$$\frac{\partial l(p|y)}{\partial y} = 2p - 3p^2, \quad \frac{\partial^2 l(p|y)}{\partial p^2} = 2 - 6p$$

$$2p - 3p^2 = 0 \quad \hat{p} \rightarrow 0 \rightarrow \text{ruled out since it does not max likelihood func.}$$

$$\hat{p} \rightarrow 2/3$$

If $\hat{p} = 0 \rightarrow$ we could not have drawn sample

$\hat{p} = 2/3$ satisfies both 1st & 2nd order conditions for a local max.

$\hat{p} = 2/3 \rightarrow$ max likelihood estimate of p

to sum up,

- ① likelihood function $l(\theta|y)$ is algebraically identical to joint pdf $f(y|\theta)$ of random sample $y_1, \dots, y_n$ where θ is a vector of unknown parameters with a true value known to fall in parameter space Ω
- ② Difference bw. $l(\theta|y)$ & $f(y|\theta)$ is that $l(\theta|y)$ is a function of θ given values of random sample y
 $f(y|\theta)$ is a function of y given particular values of θ
- ③ ML estimate of θ is that value in parameter space Ω , $\hat{\theta}$, that maximizes $\text{to max } l(\theta|y)$
- ④ Values $\hat{\theta}$ is in general a function of y that $\hat{\theta} = \hat{\theta}(y)$ so that random variable $\hat{\theta} = \hat{\theta}(y)$ where $y = (y_1, \dots, y_n)$ is ML estimator of θ
- ⑤ As a pro whether it is usual practice to maximize natural log of likelihood function rather than likelihood function
it is convenient since log-likelihood function
 $\lambda = \ln l(\theta|y)$ is composed of sums rather than products & exponential functions simplify nicely

$\ln l(\theta|y)$ & $l(\theta|y)$ attain their maximum at same value, since natural log is a monotonic function

Ex: let $Y_1, Y_2, \dots, Y_n$ be a random sample of Bernoulli random variables with

$$f(y_i | p) = p^{y_i} (1-p)^{1-y_i}, \quad i=1, \dots, n \quad 0 < p < 1$$

log likelihood function

$$\begin{aligned} L &= \ln \ell(p | y) = \ln \left[\prod_{i=1}^n f(y_i | p) \right] \\ &= \ln \left[p^{\sum y_i} (1-p)^{\sum (1-y_i)} \right] \\ &= \sum y_i \ln(p) + \sum (1-y_i) \ln(1-p) \end{aligned}$$

$$\frac{\partial L}{\partial p} = \sum y_i \frac{1}{p} - \sum (1-y_i) \frac{1}{1-p}$$

$$\frac{\partial L}{\partial p^2} = \sum y_i \frac{1}{p^2} - \left[(1-y_i) \frac{1}{(1-p)^2} \right]$$

$$\sum y_i \frac{1}{p} - \sum (1-y_i) \frac{1}{1-p} = 0$$

$$\hat{p} = \left[\sum y_i / n \right] \Rightarrow \text{sample mean}$$

$$\frac{\partial^2 L}{\partial p^2} \text{ is always negative for } 0 < p < 1$$

at least write $\frac{\partial^2 L}{\partial p^2} < 0 \Rightarrow$ higher order approximation

Ex: let $Y_1, \dots, Y_n$ be a random sample from a $N(\beta, \sigma^2)$ population. Thus unknown parameters are β & σ^2
 $(\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2))$ *estik sinav sorusu*

Find MLE of β & σ^2

$$\begin{aligned} L &= \ln \ell(\theta | y) = \ln \left\{ \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(y_i - \beta)^2}{\sigma^2}} \right\} \\ &= \ln \left\{ (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \beta)^2}{\sigma^2}} \right\} \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \beta)^2}{\sigma^2} \end{aligned}$$

$$\frac{dL}{d\beta} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i - n\beta \right)$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \beta)^2$$

$$\frac{\partial^2 L}{\partial \beta^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta)^2$$

$$\frac{\partial^2 L}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^6} \left(\sum_{i=1}^n y_i - n\beta \right)$$

$$\frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i - n\hat{\beta} \right) = 0 \Rightarrow \hat{\beta} = \sum_{i=1}^n \frac{y_i}{n} = \bar{y}$$

$$\frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \beta)^2 = 0$$

$$\frac{n}{2\sigma^2} = \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \beta)^2$$

$$\hookrightarrow \hat{\sigma}^2 = \frac{\sum (y_i - \hat{\beta})^2}{n} = \frac{\sum (y_i - \bar{y})^2}{n}$$

These unique solutions actually maximize log likelihood function can be verified by checking second order conditions for local max.

That is matrix of 2nd order conditions (derivatives) should be negative definite when evaluated at solutions of $\hat{\beta}$ & $\hat{\sigma}^2$

when evaluated at solutions of β & σ^2

$$\begin{bmatrix} \frac{\partial^2 L}{\partial \beta^2} & \frac{\partial^2 L}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \beta} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{bmatrix}$$

This is negative definite
A symmetric matrix, A , is negative