

Least Squares Estimation

Difficulty in MM & MLE is that each requires a specific assumption about distribution of random variable in question

It is possible to use estimation methods that do not require exact specification of population distribution.

↳ one method is least-squares (LS)

This method can be used to estimate moments of random variables $\mu'_r = E(y^r)$

Since math. expectation of a random variable $(y_1, \dots, y_n)$ gives moments, it is reasonable to use center of data y_i , $i=1, \dots, n$ to estimate μ'_r

One way to define center of data set is to find value that

$$\min S = \sum_{i=1}^n (y_i^r - \mu'_r)^2$$

Value of S is the sum of squared differences between y_i & expected $\mu'_r = E(y^r)$

Value of $\hat{\mu}'_r$ that min S for a given set of data values is called LS estimate of μ'_r

If $\hat{\mu}'_r$ is considered as a function of random variable $y_i \rightarrow$ it is LS estimator

Advantages of LS

- 1- It is often easier to find LS estimators since just minimizing S is simple
- 2- LS method often generates estimators with desirable properties. LS estimators are generally robust

Ex: Let $Y_1, \dots, Y_n$ be a random sample from a population with mean β & finite variance σ^2
Find LS estimator of β

$$\min S = \sum_{i=1}^n (y_i - \beta)^2$$

$$\frac{\partial S}{\partial \beta} = \sum_{i=1}^n 2(y_i - \hat{\beta}) = -2 \sum_{i=1}^n (y_i - \hat{\beta}) = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}$$

2nd order cond. is verified

Bayesian Estimation

Bayesian approach to statistics is fundamentally different from classical approach

Some aspects of this approach is quite helpful in other statistical approaches

Bayes Theorem

In classical approach parameter θ is thought to be unknown but fixed quantity

A random sample $Y_1, \dots, Y_n$ is drawn from population ideal by θ and based on observed values in some knowledge about value of θ is obtained

In bayesian approach θ is considered to be quantity whose variation can be described by probability distribution
 $\Rightarrow$ prior distribution

This is a subjective distribution based on experimenter's belief or experiences & it is formulated before data is seen

A sample is taken from a population again indexed by θ & prior distribution is updated with this sample information
Updated prior $\Rightarrow$ is called posterior distr.

This updating is done by Bayes theorem

If $B_1, B_2, \dots, B_k$ constitute a partition of sample space S , provided that each partition $P(B_i) > 0$ for $i = 1, 2, 3, \dots, k$

Therefore any event A in S , $P(A) \neq 0$

$$P(B_r | A) = \frac{P(B_r) \cdot P(A | B_r)}{\sum_{i=1}^K P(B_i) \cdot P(A | B_i)}$$

$$\begin{array}{l} \xrightarrow{P(B_1)} B_1 \xrightarrow{P(A|B_1)} A \quad P(B_1) \cdot P(A|B_1) \\ \xrightarrow{P(B_2)} B_2 \xrightarrow{P(A|B_2)} A \quad P(B_2) \cdot P(A|B_2) \\ \xrightarrow{P(B_K)} B_K \xrightarrow{P(A|B_K)} A \quad P(B_K) \cdot P(A|B_K) \end{array}$$

If we denote prior distribution by $\pi(\theta)$
 sampling distribution $f(y|\theta)$ then posterior distribution that is conditional distribution of θ given sample y ,

$$\pi(\theta | y) = \frac{f(y|\theta) \cdot \pi(\theta)}{g(y)}$$

where $f(y|\theta) \cdot \pi(\theta) \rightarrow f(y, \theta)$
 $g(y)$ is marginal distribution of y

$$g(y) = \int f(y|\theta) \cdot \pi(\theta) \cdot d\theta$$

Posterior distribution is conditional distribution
 $\hookrightarrow$ conditional upon observing sample

Posterior distribution is now used to make statements about θ , is still a random quantity
 i.e. mean of posterior distribution can be used to find a point estimate of θ

Ex: Suppose that $y_1, y_2, \dots, y_n$ is an sample of Poisson observations with mean θ . Find the MLE of θ

$$L(\theta | y) = \prod_{i=1}^n f(y_i; \theta) = \frac{\prod_{i=1}^n \theta^{y_i} e^{-\theta}}{y_i!}$$

$$= \frac{\theta^{\sum_{i=1}^n y_i} e^{-n\theta}}{\prod_{i=1}^n y_i!} \Rightarrow \ln L(\theta | y) = \sum_{i=1}^n y_i \ln \theta - n\theta - \ln(\prod_{i=1}^n y_i!)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{\sum_{i=1}^n y_i}{\theta} - n = 0$$

$$\hat{\theta} = \frac{1}{n} \sum y_i = \bar{y}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} < 0$$

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

Ex: Find method of moment estimator based on a random sample $x_1, x_2, \dots, x_n$ from following pdf

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

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

$$\mu_1 = E(x) = \int_0^1 x \theta x^{\theta-1} \cdot dx = \frac{\theta}{\theta+1}$$

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

Ex: Let $y_1, y_2, \dots, y_n$ be a random sample from $f(y; \theta) = (\theta+1) y^\theta$ where $0 < y < 1$

find method of moments estimator for θ

$$m_1 = \sum_{i=1}^n y_i = \bar{y}$$

$$\mu_1 = E(Y) = \int_0^1 y (\theta+1) y^\theta \cdot dy = \frac{\theta+1}{\theta+2}$$

$$\frac{\theta+1}{\theta+2} = \bar{y} \Rightarrow \hat{\theta} = \frac{2\bar{y}-1}{1-\bar{y}}$$

exam!!

Ex: let $x_1, \dots, x_n$ be a independent sample from a Bernoulli distribution with parameter θ

$$P(x_i = 1) = \theta, P(x_i = 0) = 1 - \theta$$

a) Find a sufficient statistics U for θ

b) Find MLE estimator of θ

c) Suppose a statistician wanted to use statistics $T(s) = (x_1 + x_2) / 2$ as an estimator of θ

Using the sufficient stat in part a, find Rao-Blackwell estimator $T(u) = E(T|U)$

d) compute MSE for T for $T(u)$ and verify (assuming $n \geq 2$) the claim of RBT in this example

a) for two samples $S_1 = (x_1, x_2, \dots, x_n)$
 $S_2 = (y_1, y_2, \dots, y_n)$

$$\left(\frac{\theta}{1-\theta} \right) \leq x_i - \leq y_i$$

which does not depend on θ iff $\sum x_i - \sum y_i$ so either $\bar{x}$ or $\sum x_i$ could be a sufficient statistic
 $\rightarrow U = \sum x_i$

$$b) \frac{\partial \ln L}{\partial \theta} = \frac{\sum x_i}{\theta} - n \frac{\sum x_i}{1-\theta} = 0$$

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

$$\frac{\partial^2 \ln L}{\partial \theta^2} < 0$$

$$c) T(s) = \frac{x_1 + x_2}{2}$$

$$T(u) = E(T|U)$$

$$E(T|U) = 1/2 E(x_1|U) + E(x_2|U) = 1/2 E(x_1|U)$$

conditional on U , sample contain exactly U

once and $n-U$ zeros. Thus $T(u) = \bar{x}$

Note that T is the mean of first 2 observations
 $T(u)$ is mean of all n observations

d) Both T & $T(u)$ are unbiased, mse are respective variances from bernoulli

$$\text{Var}(x_i) = \theta(1-\theta). \quad \text{MSE}(T) = \theta(1-\theta)/2$$

$$\text{MSE}(T(u)) = \theta(1-\theta)/n$$

$$\frac{\text{MSE}(T)}{\text{MSE}(T(u))} = n/2$$

$\text{MSE}(T) > \text{MSE}(T(u)) \rightarrow$ consistent with RBT