Chapter 2
Methods of Estimation

2.1 Introduction

In chapter one, we have discussed different optimum properties of good point estimators viz. unbiasedness, minimum variance, consistency and efficiency which are the desirable properties of a good estimator. In this chapter, we shall discuss different methods of estimating parameters which are expected to provide estimators having some of these important properties. Commonly used methods are:

1. Method of moments
2. Method of maximum likelihood
3. Method of minimum χ^2
4. Method of least squares

In general, depending on the situation and the purpose of our study we apply any one of the methods that may be suitable among the above-mentioned methods of point estimation.

2.2 Method of Moments

The method of moments, introduced by K. Pearson is one of the oldest methods of estimation. Let (X_1, X_2,…X_n) be a random sample from a population having p.d.f. (or p.m.f) $f(x,\theta)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. Further, let the first k population moments about zero exist as explicit function of θ, i.e. $\mu'_r = \mu'_r(\theta_1, \theta_2, \ldots, \theta_k)$, $r = 1, 2, \ldots, k$. In the method of moments, we equate k sample moments with the corresponding population moments. Generally, the first k moments are taken because the errors due to sampling increase with the order of the moment. Thus, we get k equations $\mu'_r(\theta_1, \theta_2, \ldots, \theta_k), = m'_r$, $r = 1, 2, \ldots, k$. Solving these equations we get the method of moment estimators (or estimates) as $m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$ (or $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$).

P.K. Sahu et al., *Estimation and Inferential Statistics*,
DOI 10.1007/978-81-322-2514-0_2

If the correspondence between μ_r' and θ is one-to-one and the inverse function is $\theta_i = f_i(\mu_1', \mu_2', \ldots, \mu_k')$, $i = 1, 2, \ldots, k$ then, the method of moment estimate becomes $\hat{\theta}_i = f_i(m_1', m_2', \ldots, m_k')$. Now, if the function $f_i()$ is continuous, then by the weak law of large numbers, the method of moment estimators will be consistent. This method gives maximum likelihood estimators when $f(x, \theta) = \exp(b_0 + b_1 x + b_2 x^2 + \ldots.)$ and so, in this case it gives efficient estimator. But the estimators obtained by this method are not generally efficient. This is one of the simplest methods. Therefore, these estimates can be used as a first approximation to get a better estimate. This method is not applicable when the theoretical moments do not exist as in the case of Cauchy distribution.

Example 2.1 Let $X_1, X_2, \ldots X_n$ be a random sample from p.d.f.

$f(x; a, b) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, 0 < x < 1; a, b > 0$. Find the estimators of a and b by the method of moments.

Solution

We know $E(x) = \mu_1^1 = \frac{a}{a+b}$ and $E(x^2) = \mu_2^1 = \frac{a(a+1)}{(a+b)(a+b+1)}$.

Hence, $\frac{a}{a+b} = \bar{x}, \frac{a(a+1)}{(a+b)(a+b+1)} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

By solving, we get $\hat{b} = \frac{(\bar{x}-1)\left(\sum x_i^2 - n\bar{x}\right)}{\sum (x_i - \bar{x})^2}$ and $\hat{a} = \frac{\bar{x}\hat{b}}{1-\bar{x}}$.

2.3 Method of Maximum Likelihood

This method of estimation is due to R.A. Fisher. It is the most important general method of estimation. Let $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ denote a random sample with joint p.d.f or p.m.f. $f\left(\underset{\sim}{x}, \theta\right), \theta \in \Theta$ (θ may be a vector). The function $f\left(\underset{\sim}{x}, \theta\right)$, considered as a function of θ, is called the likelihood function. In this case, it is denoted by $L(\theta)$. The principle of maximum likelihood consists of choosing an estimate, say $\hat{\theta}$, within the admissible range of θ, that maximizes the likelihood. $\hat{\theta}$ is called the maximum likelihood estimate (MLE) of θ. In other words, $\hat{\theta}$ will be an MLE of θ if

$$L\left(\hat{\theta}\right) \geq L(\theta) \forall\, \theta \in \Theta.$$

In practice, it is convenient to work with logarithm. Since log-function is a monotone function, $\hat{\theta}$ satisfies

$$\log L\left(\hat{\theta}\right) \geq \log L(\theta) \forall\ \theta \in \Theta.$$

Again, if $\log L(\theta)$ is differentiable within Θ and $\hat{\theta}$ is an interior point, then $\hat{\theta}$ will be the solution of

$$\frac{\partial \log \mathrm{L}(\theta)}{\partial \theta_i} = \mathrm{o},\ i = 1,\ 2, \ldots, k;\ \underset{\sim}{\theta}^{k\times 1} = (\theta_1, \theta_2, \ldots, \theta_k)'.$$

These equations are known as likelihood equations.

Problem 2.1 Let $(X_1, X_2, \ldots, X_n)$ be a random sample from $b(m, \pi)$, (m known). Show that $\hat{\pi} = \frac{1}{mn}\sum_{i=1}^{n} X_i$ is an MLE of π.

Problem 2.2 Let $(X_1, X_2, \ldots, X_n)$ be a random sample from $P\ (\lambda)$. Show that $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an MLE of λ.

Problem 2.3 Let $(X_1, X_2, \ldots, X_n)$ be a random sample from $N(\mu, \sigma^2)$. Show that $\left(\bar{X}, s^2\right)$ is an MLE of (μ, σ^2), where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$.

Example 2.2 Let $(X_1, X_2, \ldots, X_n)$ be a random sample from a population having p.d.f $f(x,\ \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty$.

Show that the sample median $\tilde{X}$ is an MLE of θ.

Answer

$$L(\theta) = \text{ Const. } e^{-\sum_{i=1}^{n} |x_i - \theta|}$$

Maximization of $L(\theta)$ is equivalent to the minimization of $\sum_{i=1}^{n} |x_i - \theta|$. Now, $\sum_{i=1}^{n} |x_i - \theta|$ will be least when $\theta = \tilde{X}$, the sample median as the mean deviation about the median is least. $\tilde{X}$ will be an MLE of θ.

Properties of MLE

(a) If a sufficient statistic exists, then the MLE will be a function of the sufficient statistic.

Proof Let T be a sufficient statistic for the family $\left\{f\left(\underset{\sim}{X}, \theta\right), \theta \in \Theta\right\}$

By the factorisation theorem, we have $\prod_{i=1}^{n} f(x_i, \theta) = g\left\{T\left(\underset{\sim}{X}\right), \theta\right\} h\left(\underset{\sim}{X}\right)$.

To find MLE, we maximize $g\left\{T\left(\underset{\sim}{x}\right), \theta\right\}$ with respect to θ. Since $g\left\{T\left(\underset{\sim}{X}\right), \theta\right\}$ is a function of θ and $\underset{\sim}{x}$ only through $T\left(\underset{\sim}{X}\right)$, the conclusion follows immediately.□

Remark 2.1 Property (a) does not imply that an MLE is itself a sufficient statistic.

Example 2.3 Let X_1, X_2,…,X_n be a random sample from a population having p.d.f.
$f\left(\underset{\sim}{X}, \theta\right) = \begin{cases} 1 & \forall\ \theta \leq x \leq \theta + 1 \\ 0 & \text{Otherwise} \end{cases}$.

Then, $L(\theta) = \begin{cases} 1 & \text{if } \theta \leq \text{Min}X_i \leq \text{Max}X_i \leq \theta + 1 \\ 0 & \text{Otherwise} \end{cases}$.

Any value of θ satisfying $\text{Max}X_i - 1 \leq \theta \leq \text{Min}X_i$ will be an MLE of θ. In particular, Min X_i is an MLE of θ, but it is not sufficient for θ. In fact, here $(\text{Min}X_i, \text{Max}X_i)$ is a sufficient statistic.

(b) If T is the MVBE, then the likelihood equation will have a solution T.

Proof Since T is an MVBE, $\dfrac{\partial \log f\left(\underset{\sim}{X}, \theta\right)}{\partial \theta} = (T - \theta)\lambda(\theta)$

Now, $\dfrac{\partial \log f\left(\underset{\sim}{X}, \theta\right)}{\partial \theta} = 0$

$$\Rightarrow \theta = T[\because \lambda(\theta) \neq 0].$$

(c) Let T be an MLE of θ and $\delta = \psi(\theta)$ be a one-to-one function of θ. Then, $d = \psi(T)$ will be an MLE of δ. □

Proof Since T is an MLE of θ, $L\left\{T\left(\underset{\sim}{X}\right)\right\} \geq L(\theta) \forall \theta$,

Since the correspondence between θ and δ is one-to-one, inverse function must exist. Suppose the inverse function is $\theta = \psi^{-1}(\delta)$.

Thus, $L(\theta) = L\{\psi^{-1}(\delta)\} = L_1(\delta)$ (say)

Now,

$$L_1(d) = L\{\psi^{-1}(d)\} = L\left[\psi^{-1}\left\{\psi\left(T\left(\underset{\sim}{X}\right)\right)\right\}\right] = L\left\{T\left(\underset{\sim}{X}\right)\right\} \geq L(\theta) = L_1(\delta).$$

Therefore, 'd' is an MLE of δ.

(d) Suppose the p.d.f. (or p.m.f.) $f(x, \theta)$ satisfies the following regularity conditions:

(i) For almost all x, $\frac{\partial f(x,\theta)}{\partial \theta}, \frac{\partial^2 f(x,\theta)}{\partial \theta^2}, \frac{\partial^3 f(x,\theta)}{\partial \theta^3}$ exists $\forall\ \theta \in \Theta$.

(ii) $\left|\frac{\partial f(x,\theta)}{\partial \theta}\right| < A_1(x)$, $\left|\frac{\partial^2 f(x,\theta)}{\partial \theta^2}\right| < A_2(x)$ and $\left|\frac{\partial^3 f(x,\theta)}{\partial \theta^3}\right| < B(x)$,
where $A_1(x)$ and $A_2(x)$ are integrable functions of x and
$\int_{-\infty}^{\infty} B(x) f(x,\ \theta) \mathrm{d}x < M$, a finite quantity

iii) $\int_{-\infty}^{\infty} \left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right)^2 f(x,\ \theta) \mathrm{d}x$ is a finite and positive quantity.

If $\hat{\theta}_n$ is an MLE of θ on the basis of a sample of size n, from a population having p.d.f. (or p.m.f.) $f(x,\theta)$ which satisfies the above regularity conditions, then

$\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ is asymptotically normal with mean '0' and variance $\left[\int_{-\infty}^{\infty}\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2 f(x,\,\theta)\mathrm{d}x\right]^{-1}$. Also, $\hat{\theta}_n$ is asymptotically efficient and consistent.

(e) An MLE may not be unique. □

Example 2.4 Let $f(x,\,\theta) = \begin{cases} 1 & \text{if } \theta \le x \le \theta + 1 \\ 0 & \text{Otherwise} \end{cases}$.

$$\text{Then, } L(\theta) = \begin{cases} 1 & \text{if } \theta \le \min\ x_i \le \max\ x_i \le \theta + 1 \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{i.e. } L(\theta) = \begin{cases} 1 & \text{if } \max\ x_i - 1 \le \theta \le \min\ x_i \\ 0 & \text{Otherwise} \end{cases}$$

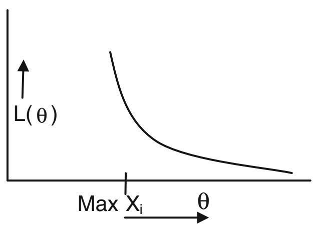
Clearly, for any value of θ, say $T_\alpha = \alpha(\mathrm{Max}x_i - 1) + (1-\alpha)\mathrm{Min}x_i,\ 0 \le \alpha \le 1$, $L(\theta)$ will be maximized. For fixed α, T_α will be an MLE. Thus, we observe that an infinite number of MLE exist in this case.

(f) An MLE may not be unbiased.

Example 2.5

$$f(x,\,\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{Otherwise} \end{cases}.$$

$$\text{Then, } L(\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \max x_i \le \theta \\ 0 & \text{Otherwise} \end{cases}.$$



From the figure, it is clear that the likelihood $L(\theta)$ will be the largest when θ = Max X_i. Therefore Max X_i will be an MLE of θ. Note that $E(\mathrm{Max}\,X_i) = \frac{n}{n+1}\theta \ne \theta$. Therefore, here MLE is a biased estimator.

(g) An MLE may be worthless.

Example 2.6

$$f(x,\pi) = \pi^x(1-\pi)^{1-x};\ x=0,1,\pi \in \left(\frac{1}{4},\frac{3}{4}\right)$$

Then, $L(\pi) = \begin{cases} \pi & \text{if } x=1 \\ 1-\pi & \text{if } x=0 \end{cases}$ i.e. $L(\pi)$ will be maximized at $\begin{cases} \pi=\frac{3}{4} & \text{if } x=1 \\ \pi=\frac{1}{4} & \text{if } x=0 \end{cases}$

Thus, $T = \frac{2X+1}{4}$ will be an MLE of θ.
Now, $E(T) = \frac{2\pi+1}{4} \neq \pi$. Thus, T is a biased estimator of π.

$$\text{MSE of } T = E(T-\pi)^2$$

$$= E\left(\frac{2x+1}{4}-\pi\right)^2 = \frac{1}{16}E\{2(x-\pi)+1-2\pi\}^2$$

$$= \frac{1}{16}E\left\{4(x-\pi)^2 + (1-2\pi)^2 + 4(x-\pi)(1-2\pi)\right\}$$

$$= \frac{1}{16}\left\{4\pi(1-\pi)+(1-2\pi)^2\right\} = \frac{1}{16}$$

Now, we consider a trivial estimator $\delta(x) = \frac{1}{2}$.
MSE of $\delta(x) = \left(\frac{1}{2}-\pi\right)^2 \leq \frac{1}{16}$ = MSE of T $\forall \pi \in \left(\frac{1}{4},\frac{3}{4}\right)$
Thus, in the sense of mean square error MLE is meaningless.
(h) An MLE may not be consistent

Example 2.7

$$f(x,\ \theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{if } \theta \text{ is rational} \\ (1-\theta)^x\theta^{1-x} & \text{if } \theta \text{ is rational} \quad 0<\theta<1, x=0,1. \end{cases}$$

An MLE of θ is $\hat{\theta}_n = \bar{X}$. Here, $\hat{\theta}_n$ is not a consistent estimator of θ.
(i) The regularity conditions in (d) are not necessary conditions.

Example 2.8

$$f(x,\theta) = \frac{1}{2}e^{-|x-\theta|},\quad \begin{matrix} -\infty<x<\infty \\ -\infty<\theta<\infty \end{matrix}$$

Here, regularity conditions do not hold. However, the MLE (=sample median) is asymptotically normal and efficient.

Example 2.9 Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x,\alpha,\beta) = \beta e^{-\beta(x-\alpha)};\ \alpha \leq x < \infty$ and $\beta > 0$.
Find MLE's of α, β.

Solution

$$L(\alpha, \beta) = \beta^n e^{-\beta \sum_{i=1}^{n} (x_i - \alpha)}$$

$$\log_e L(\alpha, \beta) = n \log_e \beta - \beta \sum_{i=1}^{n} (x_i - \alpha)$$

$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} - \sum (x_i - \alpha)$ and $\frac{\partial \log L}{\partial \alpha} = n\beta$.

Now, $\frac{\partial \log L}{\partial \alpha} = 0$ gives us $\beta = 0$ which is nonadmissible. Thus, the method of differentiation fails here.

Now, from the expression of $L(\alpha, \beta)$, it is clear that for fixed $\beta(>0)$, $L(\alpha, \beta)$ becomes maximum when α is the largest. The largest possible value of α is $X_{(1)}$ = Min x_i.

Now, we maximize $L\{X_{(1)}, \beta\}$ with respect to β. This can be done by considering the method of differentiation.

$$\frac{\partial \log L\{x_{(1)}, \beta\}}{\partial \beta} = 0 \Rightarrow \frac{n}{\beta} - \sum (x_i - \min x_i) = 0 \quad \Rightarrow \beta = \frac{n}{\sum (x_i - \min x_i)}$$

So, the MLE of (α, β) is $\left\{\min x_i, \frac{n}{\sum (x_i - \min x_i)}\right\}$.

Example 2.10 Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x, \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha \le x \le \beta \\ 0, & \text{Otherwise} \end{cases}$

(a) Show that the MLE of (α, β) is (Min X_i, Max X_i).

(b) Also find the estimators of α and β by the method of moments.

Proof

$$\text{(a)} L(\alpha, \beta) = \frac{1}{(\beta - \alpha)^n} \quad \text{if } \alpha \le \text{Min} x_i < \text{Max} x_i \le \beta \tag{2.1}$$

It is evident from (2.1), that the likelihood will be made as large as possible when $(\beta - \alpha)$ is made as small as possible. Clearly, α cannot be larger than Min x_i and β cannot be smaller than Max x_i; hence, the smallest possible value of $(\beta - \alpha)$ is (Max x_i − Min x_i). Then the MLE'S of α and β are $\hat{\alpha} = \text{Min}\, x_i$ and $\hat{\beta} = \text{Max}\, x_i$, respectively.

(b) We know $E(x) = \mu_1^1 = \frac{\alpha + \beta}{2}$ and $V(x) = \mu_2 = \frac{(\beta - \alpha)^2}{12}$

Hence, $\frac{\alpha + \beta}{2} = \bar{x}$ and $\frac{(\beta - \alpha)^2}{12} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

By solving, we get $\hat{\alpha} = \bar{x} - \sqrt{\frac{3\sum(x_i-\bar{x})^2}{n}}$ and $\hat{\beta} = \bar{x} + \sqrt{\frac{3\sum(x_i-\bar{x})^2}{n}}$

Successive approximation for the estimation of parameter

It frequently happens that the likelihood equation is by no means easy to solve. A general method in such cases is to assume a trial solution and correct it by an extra term to get a more accurate solution. This process can be repeated until we get the solution to a sufficient degree of accuracy.

Let L denote the likelihood and θ^* be the MLE.

Then $\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta^*} = 0$. Suppose θ_0 is a trial solution of $\frac{\partial \log L}{\partial \theta} = 0$

Then

$0 = \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta^*} = \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta_0} + (\theta^* - \theta_0)\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta=\theta_0}$ + terms involving $(\theta^* - \theta_0)$ with powers higher than unity.

$\Rightarrow 0 \simeq \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta_0} + (\theta^* - \theta_0)\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta=\theta_0}$, neglecting the terms involving $(\theta^* - \theta_0)$ with powers higher than unity.

$\Rightarrow 0 \simeq \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta_0} - (\theta^* - \theta_0)\mathrm{I}(\theta_0)$, where $\mathrm{I}(\theta) = -E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)$.

Thus, the first approximate value of θ is

$$\theta^{(1)} = \theta_0 + \left\{ \frac{\left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\theta_0}}{\mathrm{I}(\theta_0)} \right\}.$$

Example 2.11 Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x, \theta) = \frac{1}{\pi\left\{1 + (x-\theta)^2\right\}}$

Here, $\frac{\partial \log f(x,\theta)}{\partial \theta} = \frac{2(x-\theta)}{1+(x-\theta)^2}$; and so the likelihood equation is $\sum_{i=1}^{n} \frac{(x_i-\theta)}{1+(x_i-\theta)^2} = 0$; clearly it is difficult to solve for θ.

So, we consider successive approximation method.

In this case, $\mathrm{I}(\theta) = \frac{n}{2}$.

Here, the first approximation is $\theta^{(1)} = \theta_0 + \frac{4}{n}\sum_{i=1}^{n} \frac{(x_i-\theta_0)}{1+(x_i-\theta_0)^2}$,

θ_0 being a trial solution.

Usually, we take θ_0 = sample median.

2.4 Method of Minimum χ^2

This method may be used when the population is grouped into a number of mutually exclusive and exhaustive class and the observations are given in the form of frequencies.

Suppose there are k classes and $\pi_i(\theta)$ is the probability of an individual belonging to the ith class. Let f_i denote the sample frequency. Clearly, $\sum_{i=1}^{k} \pi_i(\theta) = 1$ and $\sum_{i=1}^{k} f_i = n$.

The discrepancy between observed frequency and the corresponding expected frequency is measured by the Pearsonian χ^2, which is given by $\chi^2 = \sum_{i=1}^{k} \frac{\{f_i - n\pi_i(\theta)\}^2}{n\pi_i(\theta)} = \sum \frac{f_i^2}{n\pi_i(\theta)} - n$.

The principle of the method of minimum χ^2 consists of choosing an estimate of θ, say $\hat{\theta}$, we first consider the minimum χ^2 equations $\frac{\partial \chi^2}{\partial \theta_i} = 0$, $i = 1, 2, \ldots, r$ and $\theta_i =$ ith component of θ.

It can be shown that for large n, the min χ^2 equations and the likelihood equations are identical and provides identical estimates.

The method of minimum χ^2, is found to be more troublesome to apply in many cases, and has no improvement on the maximum likelihood method. This method can be used when maximum likelihood equations are difficult to solve. In particular situations, this method may be simple. To avoid the difficulty in minimum χ^2 method, we consider another measure of discrepancy, which is given by $\chi'^2 = \sum_{i=1}^{k} \frac{\{f_i - n\pi_i(\theta)\}^2}{f_i}$, χ'^2 is called modified Pearsonian χ^2. Now, we minimize, instead of χ^2, with respect to θ.

It can be shown that for large n the estimates obtained by min χ^2 would also be approximately equal to the MLE's. Difficulty arises if some of the classes are empty. In this case, we minimize

$$\chi''^2 = \sum_{i:f_i \neq 0} \frac{\{f_i - n\pi_{\mathrm{i}}(\theta)\}^2}{f_i} + 2\mathrm{M},$$

where M = sum of the expected frequencies of the empty classes.

Example 2.12 Let $(x_1, x_2, \ldots, x_n)$ be a given sample of size n. It is to be tested whether the sample comes from some Poisson distribution with unknown mean μ. How do you estimate μ by the method of modified minimum chi-square?

Solution

Let $x_1, x_2, \ldots, x_n$ be arranged in k groups such that there are

n_i observations with $x = i, i = r+1, \ldots, r+k-2$

n_L observations $x \leq r$

n_u observations with $x \geq r + k - 1$ so that the smallest and the largest values of x, which are fewer, are pooled together and$n_L + \sum_{i=r+1}^{r+k-2} n_i + n_u = n$.

Let $\pi_i(\mu) = P(x = i) = \frac{e^{-\mu}\mu^i}{i!}$, $\pi_L(\mu) = P(x \leq r) = \sum_{i=0}^{r} \pi_i(\mu)$ and $\pi_u(\mu) = P(x \geq r + k - 1) = \sum_{i=r+k-1}^{\infty} \pi_i(\mu)$.

Now using $\sum_{i=1}^{k} \frac{n_i}{\pi_i(\theta)} \frac{\partial \pi_i(\theta)}{\partial \theta_j} = 0$, $j = 1, 2, \ldots p$ we have $n_L \frac{\sum_{i=0}^{r} \left(\frac{i}{\mu} - 1\right)\pi_i(\mu)}{\sum_{i=0}^{r} \pi_i(\mu)} + \sum_{i=r+1}^{r+k-2} n_i \left(\frac{i}{\mu} - 1\right) + n_u \frac{\sum_{i=r+k-1}^{\infty} \left(\frac{i}{\mu} - 1\right)\pi_i(\mu)}{\sum_{i=r+k-1}^{\infty} \pi_i(\mu)} = 0$.

Since there is only one parameter, i.e. $p = 1$ we get the only above equation. By solving, we get

$$n\hat{\mu} = n_L \frac{\sum_{i=0}^{r} i\pi_i(\mu)}{\sum_{i=0}^{r} \pi_i(\mu)} + \sum_{i=r+1}^{r+k-2} in_i + n_u \frac{\sum_{i=r+k-1}^{\infty} i\pi_i(\mu)}{\sum_{i=r+k-1}^{\infty} \pi_i(\mu)}$$

= sum of all x's

Hence, $\hat{\mu}$ is approximately the sample mean $\bar{x}$.

2.5 Method of Least Square

In the method of least square , we consider the estimation of parameters using some specified form of the expectation and second moment of the observations. For fitting a curve of the form $y = f(x, \beta_0, \beta_1, \ldots, \beta_p)$ to the data (x_i, y_i), i = 1, 2,…n, we may use the method of least squares. This method consists of minimizing the sum of squares.

$S = \sum_{i=1}^{n} \varepsilon_i^2$, where $\varepsilon_i = y_i - f(x_i, \beta_0, \beta_1, \ldots, \beta_p)$, i = 1, 2,…,n with respect to $\beta_0, \beta_1, \ldots, \beta_p$. Sometimes, we minimize $\sum w_i \varepsilon_i^2$ instead of $\sum \varepsilon_i^2$. In that case, it is called a weighted least square method.

To minimize S, we consider (p + 1) first order partial derivatives and get (p + 1) equations in (p + 1) unknowns. Solving these equations, we get the least square estimates of β_i's.

In general, the least square estimates do not have any optimum properties even asymptotically. However, in case of linear estimation this method provides good estimators. When $f(x_i, \beta_0, \beta_1, \ldots, \beta_p)$ is a linear function of the parameters and the

x-values are known, least square estimators will be BLUE. Again, if we assume that ε_i's are independently and identically normally distributed, then a linear estimator of the form $\underset{\sim}{a}' \underset{\sim}{\beta}$ will be MVUE for the entire class of unbiased estimators. In general, we consider n uncorrelated observations $y_1, y_2, \ldots y_n$ such that $E(y_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$.

$$V(y_i) = \sigma^2,\ i = 1, 2, \ldots\ldots, n,\ x_{1i} = 1 \forall i,$$

where $\beta_1, \beta_2 \ldots\ldots\ldots \beta_k$ and σ^2 are unknown parameters. If Y and β^* stand for column vectors of the variables y_i and parameters β_j and if $X = (x_{ji})$ be an $(n \times k)$ matrix of known coefficients x_{ji} then the above equation can be written as

$$E(Y) = X\beta^*$$

$$V(e) = E(ee') = \sigma^2 I$$

where $e = Y - X\beta^*$ is an $(n \times 1)$ vector of error random variable with $E(e) = 0$ and I is an $(n \times n)$ identity matrix. The least square method requires that β's be such calculated that $\phi = e'e = (Y - X\beta^*)'(Y - X\beta^*)$ be the minimum. This is satisfied when

$$\frac{\partial \phi}{\partial \beta^*} = 0$$

$$\text{Or, } 2X'(Y - X\beta^*) = 0.$$

The least square estimators β's is thus given by the vector $\hat{\beta}^* = (X'X)^{-1} X'Y$.

Example 2.13 Let $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$, $i = 1, 2, \ldots\ldots, n$ or $E(y_i) = \beta_1 x_{1i} + \beta_2 x_{2i},\ x_{1i} = 1$ for all i.

Find the least square estimates of β_1 and β_2. Prove that the method of maximum likelihood and the method of least square are identical for the case of normal distribution.

Solution

In matrix notation we have

$$E(Y) = X\beta^*, \text{ where } X = \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{2n} \end{pmatrix},\ \beta^* = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Now,

$$\hat{\beta}^* = (X'X)^{-1}X'Y$$

Here, $X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{21} & x_{22} & \dots & x_{2n} \end{pmatrix} \begin{pmatrix} 1 & x_{21} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{2n} \end{pmatrix} = \begin{pmatrix} n & \sum x_{2i} \\ \sum x_{2i} & \sum x_{2i}^2 \end{pmatrix}$

$$X'Y = \begin{pmatrix} \sum y_i \\ \sum x_{2i} y_i \end{pmatrix}$$

$$\begin{aligned} \therefore \hat{\beta}^* &= \frac{1}{n\sum x_{2i}^2 - (\sum x_{2i})^2} \begin{pmatrix} \sum x_{2i}^2 & -\sum x_{2i} \\ -\sum x_{2i} & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_{2i} y_i \end{pmatrix} \\ &= \frac{1}{n\sum x_{2i}^2 - (\sum x_{2i})^2} \begin{pmatrix} \sum x_{2i}^2 \sum y_i - \sum x_{2i} \sum x_{2i} y_i \\ -\sum x_{2i} \sum y_i + n \sum x_{2i} \sum y_i \end{pmatrix} \end{aligned}$$

Hence,

$$\begin{aligned} \hat{\beta}_2 &= \frac{n\sum x_{2i} \sum y_i - \sum x_{2i} \sum y_i}{n\sum x_{2i}^2 - (\sum x_{2i})^2} = \frac{\sum x_{2i} \sum y_i - n\bar{x}_2\bar{y}}{\sum x_{2i}^2 - n\bar{x}_2^2} \\ &= \frac{\sum (x_{2i} - \bar{x}_2)(y_i - \bar{y})}{\sum (x_{2i} - \bar{x}_2)^2} \end{aligned}$$

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum x_{2i}^2 \sum y_i - \sum x_{2i} \sum x_{2i} y_i}{n\sum x_{2i}^2 - (\sum x_{2i})^2} \\ &= \frac{\bar{y} \sum x_{2i}^2 - \bar{x}_2 \sum x_{2i} y_i}{\sum x_{2i}^2 - n\bar{x}^2} \\ &= \bar{y} + \frac{\bar{y} n \bar{x}_2^2 - \bar{x}_2 \sum x_{2i} y_i}{\sum x_{2i}^2 - n\bar{x}_2^2} \\ &= \bar{y} - \bar{x}_2 \hat{\beta}_2 \end{aligned}$$

Let y_i be an independent $N(\beta_1 + \beta_2 x_i, \sigma^2)$ variate, $i = 1, 2, \ldots\ldots, n$ so that $E(y_i) = \beta_1 + \beta_2 x_i$. The estimators of β_1 and β_2 are obtained by the method of least square on minimizing

$$\phi = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$$

The likelihood estimate is

$$L = \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n e^{-\frac{1}{2\sigma^2}\sum (y_i - \beta_1 - \beta_2 x_i)^2}$$

L is maximum when $\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$ is minimum. By the method of maximum likelihood, we choose β_1 and β_2 such that $\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2 = \phi$ is minimum. Hence, both the methods of least square and maximum likelihood estimator are identical.

Example 2.14 Let $X_1, X_2, \ldots X_n$ be a random sample from p.d.f. $f(x; \theta, r) = \frac{1}{\theta^r \Gamma(r)} e^{-x/\theta} x^{r-1},\ x > 0;\ \theta > 0,\ r > 0.$

Find estimator of θ and r by

(i) Method of moments
(ii) Method of maximum likelihood

Answer

(i) Here, $E(x) = \mu_1^1 = r\theta,\ E(x^2) = \mu_2^1 = r(r+1)\theta^2$

$$m_1^1 = \bar{x},\ m_2^1 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Hence, $r\theta = \bar{x},\ r(r+1)\theta^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

By solving, we get $\hat{r} = \frac{n\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ and $\hat{\theta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n\bar{x}}$

(ii) $L = \frac{1}{\theta^{nr} (\Gamma(r))^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} x_i^{r-1}$

$\log L = -nr \log\theta - n\log\Gamma(r) - \frac{1}{\theta}\sum_{i}^{n} x_i + (r-1)\sum_{i=1}^{n} \log x_i$

Now, $\frac{\partial \log L}{\partial\theta} = -\frac{nr}{\theta} + \frac{n\bar{x}}{\theta^2} = 0 \Rightarrow \hat{\theta} = \frac{\bar{x}}{r}$

$$\frac{\partial \log L}{\partial r} = -n\log\theta - n\frac{\partial \log \Gamma(r)}{\partial r} + \sum_{i=1}^{n} \log x_i$$
$$= n\log r - n\frac{\Gamma'(r)}{\Gamma(r)} - n\log \bar{x} + \sum_{i}^{n} \log x_i$$

It is, however, difficult to solve the equation $\frac{\partial \log L}{\partial r} = 0$ and to get the estimate of r. Thus, for this example estimators of θ and r are more easily obtained by the method of moments than the method of maximum likelihood.

Example 2.15 If a sample of size one is drawn from the p.d.f $f(x,\beta) = \frac{2}{\beta^2}(\beta - x), 0 < x < \beta$.

Find $\hat{\beta}$, the MLE of β and β^*, the estimator of β based on method of moments. Show that $\hat{\beta}$ is biased, but β^* is unbiased. Show that the efficiency of $\hat{\beta}$ w.r.t. β^* is 2/3.

Solution

$$L = \frac{2}{\beta^2}(\beta - x)$$

$$\log L = \log 2 - 2\log\beta + \log(\beta - x)$$

$$\frac{\partial \log L}{\partial \beta} = -\frac{2}{\beta} + \frac{1}{\beta - x} = 0 \Rightarrow \beta = 2x$$

Thus, the MLE of β is given by $\widehat{\beta} = 2x$.
Now, $E(x) = \frac{2}{\beta^2}\int_0^{\beta}(\beta x - x^2)\mathrm{d}x = \frac{\beta}{3}$
Hence, $\frac{\beta}{3} = x \Rightarrow \beta = 3x$
Thus, the estimator of β based on method of moment is given by $\beta^* = 3x$. Now,

$$E\left(\hat{\beta}\right) = 2 \times \frac{\beta}{3} = \frac{2\beta}{3} \neq \beta$$
$$E(\beta^*) = 3 \times \frac{\beta}{3} = \beta$$

Hence, $\hat{\beta}$ is biased but β^* is unbiased.

Again,

$$E\left(x^2\right) = \frac{2}{\beta^2}\int_0^\beta \left(\beta x^2 - x^3\right)\mathrm{d}x = \frac{\beta^2}{6}$$

$$\begin{aligned}
\therefore V(x) &= \frac{\beta^2}{6} - \frac{\beta^2}{9} = \frac{\beta^2}{18} \\
V(\beta^*) &= 9V(x) = \frac{\beta^2}{2} \\
V\left(\hat{\beta}\right) &= 4V(x) = \frac{2}{9}\beta^2 \\
M\left(\hat{\beta}\right) &= V\left(\hat{\beta}\right) + \left[E\left(\hat{\beta}\right) - \beta\right]^2 \\
&= \frac{2}{9}\beta^2 + \left(\frac{2}{3}\beta - \beta\right)^2 \\
&= \frac{1}{3}\beta^2
\end{aligned}$$

Thus, the efficiency of $\hat{\beta}$ with respect to β^* is $\frac{2}{3}$.

http://www.springer.com/978-81-322-2513-3

Estimation and Inferential Statistics
Sahu, P.K.; Pal, S.R.; Das, A.K.
2015, XXIV, 317 p. 23 illus., Hardcover
ISBN: 978-81-322-2513-3