UNIT 6 POINT ESTIMATION

Structure

6.1 INTRODUCTION

In previous unit, we have discussed some important properties of an estimator such as unbiasedness, consistency, efficiency, sufficiency, etc. And according to Prof. Ronald A. Fisher, if an estimator possess these properties then it is said to be a good estimator. Now, our point is to search such estimators which possess as many of these properties as possible. In this unit, we shall discuss some frequently used methods of finding point estimate such as method of maximum likelihood, method of moments and method of least squares.

This unit is divided into seven sections. Section 6.1 is introductory in nature. The point estimation and frequently used methods of point estimation are explored in Section 6.2. The most important method of point estimation i.e. method of maximum likelihood and the properties of its estimators are described in Section 6.3. The method of moments with properties and drawbacks of moment estimators are described in Section 6.4. Section 6.5 is devoted to the method of least squares and its properties. Unit ends by providing summary of what we have discussed in this unit in Section 6.6 and solution of exercises in Section 6.7.

Objectives

After going through this unit, you should be able to:

- define and obtain the point estimation;
- define and obtain the likelihood function;
- explore the different methods of point estimation;
- explain the method of maximum likelihood;
- describe the properties of maximum likelihood estimators;
- discuss the method of moments;
- describe the properties of moment estimators;
- explain the method of least squares; and
- explore the properties of least squares estimators.

6.2 POINT ESTIMATION

The technique of estimating the unknown parameter with a single value is known as point estimation.

There are so many situations in our day to day life where we need to estimate the some unknown parameter(s) of the population on the basis on the sample observations. For example, a house wife may want to estimate the monthly expenditure, a sweet shopkeeper may want to estimate the sale of sweets on a day, a student may want to estimates the study hours for reading of a particular unit of this course, etc. This need is fulfilled by the technique of estimation. So the technique of finding an estimator to produce an estimate of the unknown parameter is called estimation.

We have already said that estimation is broadly divided into two categories namely:

- Point estimation and
- Interval estimation

If, we find a single value with the help of sample observations which is taken as the estimated value of unknown parameter then this value is known as point estimate and the technique of estimating the unknown parameter with a single value is known as **"point estimation"**.

If instead of finding a single value to estimate the unknown parameter if we find two values between which the parameter may be considered to lie with certain probability(confidence) is known as interval estimate of the parameter and this technique of estimating is known as **"interval estimation"**. For example, if we estimate the average weight of men living in a colony on the basis of sample mean, say, 62 kg then 62 kg is called point estimate of average weight of men in the colony and this procedure is called as point estimation. If we estimate the average weight of men by an interval, say, [40,110] with 90% confidence that true value of the weight lie in this interval then this interval is called interval estimate and this procedure is called as interval estimation.

Now, the question may arise in your mind that "how point and interval estimates are obtained?" So we will describe some of the important and frequently used methods of point estimation in the subsequent sections of this unit and methods of interval estimation in the next unit.

6.2.1 Methods of Point Estimation

Some of the important and frequently used methods of point estimation are:

1. Method of maximum likelihood
2. Method of moments
3. Method of least squares
4. Method of minimum chi-square
5. Method of minimum variance

The method of maximum likelihood, method of moments and method of least squares will be discussed in detail in subsequent sections one by one and other methods are beyond the of scope of this course.

Now, try the following exercises.

E1) Find which technique of estimation (point estimation or interval estimation) is used in each case given below:

(i) An investigator estimates average income Rs. 1.5 lack per annum of the people living in a particular geographical area, on the basis of a sample of 50 people taken from that geographical area.

(ii) A product manager of a company estimates the average life of electric bulbs in the range 800 hours and 1000 hours, with certain confidence, on the basis of a sample of 20 bulbs.

(iii) A pathologist estimates the mean time required to complete a certain analysis in the range 30 minutes to 45 minutes, with certain confidence, on the basis of a random sample of 25.

E2) List any three methods of point estimation.

6.3 METHOD OF MAXIMUM LIKELIHOOD

For describing the method of maximum likelihood, first we have to define likelihood function.

Likelihood Function

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population with joint probability density (mass) function $f(x_1, x_1, ..., x_n, \theta)$ of sample values then likelihood function is denoted by L(θ) and is defined as follows:

$$L(\theta) = f(x_1, x_1, ..., x_n, \theta)$$

For discrete case,

$$L(\theta) = P[X_1 = x_1]P[X_2 = x_2]...P[X_n = x_n]$$

For continuous case,

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

The main difference between the joint probability density (mass) function and the likelihood function is that in joint probability density (mass) function we consider the X's as variables and the parameter θ as fixed and consider as a function of sample observations while in the likelihood function we consider the parameter θ as the variable and the X's as fixed and consider as a function of parameter θ.

The process of finding the likelihood function is described by taking an example.

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from exponential distribution (θ) whose pdf is given by

$$f(x, \theta) = \theta e^{-\theta x}; \qquad x > 0, \lambda > 0$$

Then the likelihood function of parameter θ can be obtained as

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

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

$$= \theta^{\overbrace{1+1+...+1}^{n-\text{times}}} e^{-\theta(x_1+x_2+...+x_n)}$$

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

The likelihood principal states that all the information in a sample to draw the inference about the value of unknown parameter θ is found in the corresponding likelihood function. Therefore, the likelihood function gives the relative likelihoods for different values of the parameters, given the sample data.

From theoretical point of view, one of the most important methods of point estimation is method of maximum likelihood because it generally gives very good estimators as judged from various criteria. It was initially given by Prof. C.F. Gauss but later on it was used as a general method of estimation by Prof. Ronald A. Fisher in 1912. The principal of maximum likelihood estimation is to find /estimate /choose the value of unknown parameter which would most likely generate the observed data. We know that the likelihood function gives the relative likelihoods for different values of the parameters for the observed data. Therefore, we search the value of unknown parameter for which the likelihood function is maximum corresponding to the observed data. The concept of maximum likelihood estimation is explained with a simple example given below:

Suppose, we toss a coin 5 times and we observe 3 heads and 2 tails. Instead of assuming that the probability of getting head is p = 0.5, we want to find / estimate the value of p that makes the observed data most likely. Since number of heads follows the binomial distribution, therefore, the probability (likelihood function) of getting 3 heads in 5 tosses is given by

$$P\left[X = 3\right] = {}^{5}C_{3}(p)^{3}\left(1-p\right)^{2}$$

Imagine that p was 0.1 then

$$P\left[X = 3\right] = {}^{5}C_{3}(0.1)^{3}\left(0.1\right)^{2} = 0.0081$$

Similarly, for different values of p the probability of getting 3 heads in 5 tosses is given in Table 6.1 given below:

Table 6.1: Probability/Likelihood Function Corresponding to Different Values of p

S. No.	p	Probability/ Likelihood Function
1	0.1	0.0081
2	0.2	0.0512
3	0.3	0.1323
4	0.4	0.2304
5	0.5	0.3125
6	0.6	0.3456
7	0.7	0.3087
8	0.8	0.2048
9	0.9	0.0729

From Table 6.1, we can conclude that p is more likely to be 0.6 because at p = 0.6 the probability is maximum or the likelihood function is maximum.

Therefore, principle of maximum likelihood (ML) consists in finding an estimate for the unknown parameter θ within the admissible range of θ, i.e. within parameter space Θ, which makes the likelihood function as large as possible, that is, maximize the likelihood function. Such an estimate is known as maximum likelihood estimate for unknown parameter θ. Thus, if there exists an estimate, say, $\hat{\theta}(x_1, x_2, ..., x_n)$ of the sample values which maximizes likelihood function L(θ) is known as **"maximum likelihood estimate"**. That is,

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$$

^ - is used to represent the estimate / estimator of parameter and read as cap i.e.
$\hat{\theta}$ - means it is the estimate of parameter θ and read as theta cap.

For maximising the likelihood function, the theory of maxima or minima (discussed in Unit 6 of MST-002) is applied i.e. we differentiate L(θ) or L partially with respect to the parameter θ and put it equal to zero. The equation so obtain is known as likelihood equation, that is,

$$\frac{\partial}{\partial \theta} L = 0$$

Then we solve the likelihood equation for parameter θ which gives the ML estimate if its second derivative is negative at $\theta = \hat{\theta}$, that is,

$$\left.\frac{\partial^2}{\partial \theta^2} L\right|_{\theta=\hat{\theta}} < 0$$

Since likelihood function (L) is the product of n functions so differentiating L as such is very difficult. Also L is always non-negative and log L remains a finite value. So log L attains maximum when L is maximum. Hence, we can consider log L in place of L and we find ML estimate as

$$\frac{\partial}{\partial \theta}(\log L) = 0$$

provided,

$$\left.\frac{\partial^2}{\partial \theta^2}(\log L)\right|_{\theta=\hat{\theta}} < 0$$

When there are more than one parameter, say, $\theta_1, \theta_2, \ldots, \theta_k$ then ML estimates of these parameters can be obtained as the solution of k simultaneous likelihood equations

$$\frac{\partial}{\partial \theta_i}(\log L) = 0; \qquad \text{for all } i = 1, 2, ..., k$$

provided, the matrix of derivatives

$$\left.\frac{\partial^2}{\partial \theta_i \partial \theta_j}(\log L)\right|_{\theta_i=\hat{\theta}_i \;\&\; \theta_j=\hat{\theta}_j} < 0; \quad \text{for all } i \neq j = 1, 2, ..., k$$

Let us explain the procedure of ML estimation with the help of some examples.

Example 1: If the number of weekly accidents occurring on a mile stretch of a particular road follows Poisson distribution with parameter λ then find the maximum likelihood estimate of parameter λ on the basis of the following data:

Number of Accidents	0	1	2	3	4	5	6
Frequency	10	12	12	9	5	3	1

Solution: Here, the number of weekly accidents occurring on a mile stretch of a particular road follows Poisson distribution with parameter λ so pmf of the Poisson distribution is given by

$$P[X = x] = \frac{e^{-\lambda}\lambda^{x}}{x!}; \qquad x = 0, 1, 2, \ldots \ \& \lambda > 0$$

First we find the theoretical ML estimate of the parameter λ as:

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from this Poisson distribution, therefore, likelihood function of parameter λ can be obtained as

$$L(\lambda) = L = P[X = x_1].P[X = x_2] \ldots P[X = x_n]$$

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} . \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \ldots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= \frac{e^{\underbrace{-\lambda-\lambda-\ldots-\lambda}_{n-\text{times}}} \lambda^{x_1+x_2+\ldots+x_n}}{x_1!\,x_2!\ldots x_n!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \qquad \ldots (1)$$

Taking log on both sides

$$\log L = -n\lambda + \sum_{i=1}^{n} x_i \log\lambda - \log\prod_{i=1}^{n} x_i! \qquad \ldots (2)$$

Differentiating equation (2) partially with respect to λ, we get

$$\frac{\partial}{\partial\lambda}(\log L) = -n + \sum_{i=1}^{n} x_i\left(\frac{1}{\lambda}\right) - 0 \qquad \ldots (3)$$

For maxima or minima, that is, for finding ML estimate, we put

$$\frac{\partial}{\partial\lambda}(\log L) = 0$$

$$\Rightarrow -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

Now, we obtain the second derivative, that is, differentiating equation (3) partially with respect to λ, we get

$$\frac{\partial^2}{\partial\lambda^2}(\log L) = 0 + \sum_{i=1}^{n} x_i\left(-\frac{1}{\lambda^2}\right)$$

Put $\lambda = \overline{x}$, we have

$$\left.\frac{\partial^2}{\partial\lambda^2}(\log L)\right|_{\lambda=\overline{x}} = -\sum_{i=1}^{n} x_i\left(\frac{1}{\overline{x}^2}\right) < 0 \quad \text{for all values of } x_i\text{'s}$$

Therefore, maximum likelihood estimate of parameter λ of Poisson distribution is sample mean.

Since we observed that the sample mean is the maximum likelihood estimate of parameter λ so we calculate the sample mean of the given data as

S. No.	Number of Accidents(X)	Frequency(f)	fX
1	0	10	0
2	1	12	12
3	2	12	24
4	3	9	27
5	4	5	20
6	5	3	15
7	6	1	6
		N = 52	$\sum fX = 104$

The formula for calculating mean is

$$\bar{X} = \frac{1}{N}\sum fX \quad \text{where, N is the total number of accidents}$$

$$= \frac{1}{52} \times 104 = 2$$

Hence, maximum likelihood estimate of λ is 2.

Example 2: For random sampling from normal population N(μ, σ^2), find the maximum likelihood estimators for μ and σ^2.

Solution: Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population N(μ, σ^2), whose probability density function is given by

$$f\left(x, \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

Therefore, the likelihood function for parameters μ and σ^2 can be obtained as

$$L\left(\mu, \sigma^2\right) = L = f\left(x_1, \mu, \sigma^2\right).f\left(x_2, \mu, \sigma^2\right) ... f\left(x_n, \mu, \sigma^2\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_1-\mu)^2} . \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_2-\mu)^2} ... \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \quad \text{... (4)}$$

Taking log on both sides of equation (4), we get

$$\log L = \frac{n}{2}\log 1 - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 \quad \text{... (5)}$$

Differentiating equation (5) partially with respect to μ and σ^2 respectively, we get

$$\frac{\partial}{\partial\mu}(\log L) = -0 - 0 - 0 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}2\left(x_i - \mu\right)(-1) \quad \text{... (6)}$$

$$\frac{\partial}{\partial\sigma^2}(\log L) = -0 - 0 - \frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2(\sigma^2)^2}(-1)\sum_{i=1}^{n}\left(x_i - \mu\right)^2 \quad \text{... (7)}$$

For finding ML estimate of μ, we put

$$\frac{\partial}{\partial \mu}\left(\log L\right)=0$$

$$\Rightarrow -\frac{1}{2\sigma^2}\sum_{i=1}^{n} 2\left(x_i-\mu\right)\left(-1\right)=0$$

$$\Rightarrow \sum_{i=1}^{n}\left(x_i-\mu\right)=0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0 \Rightarrow \hat{\mu}=\frac{1}{n}\sum_{i=1}^{n} x_i=\overline{x}$$

Thus, the ML estimate for μ is the observed sample mean $\overline{x}$.

For ML estimate of σ^2, we put

$$\frac{\partial}{\partial \sigma^2}\log L=0$$

$$\Rightarrow -\frac{n}{2\sigma^2}-(-1)\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2=0$$

$$\Rightarrow \frac{-n\sigma^2+\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{2\sigma^4}=0$$

$$\Rightarrow -n\sigma^2+\sum_{i=1}^{n}\left(x_i-\mu\right)^2=0$$

$$\Rightarrow \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$$

$$\Rightarrow \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2=s'^2$$

Thus, the ML estimates for μ and σ^2 are $\overline{x}$ and s'^2 respectively.
Hence, ML estimators for μ and σ^2 are $\overline{X}$ and S'^2 respectively.

Note 1: Since throughout the course we are using capital letters for estimators therefore in the last line of above example we use capital letters for ML estimators for μ and σ^2.

Note 2: Here, the maxima and minima method is used to obtain the ML estimates when the range of random variable is independent of parameter θ. Whereas when the range of random variable is involved or depends on parameter θ then this method fails to find the ML estimates. In such cases, we use order statistics to maximize the likelihood function. Following example will explain this concept.

Example 3: Obtain the ML estimators for α and β for the uniform or rectangular population whose pdf is given by

$$f\left(x,\alpha,\beta\right)=\begin{cases}\dfrac{1}{\beta-\alpha}; & \alpha\le x\le\beta\\ 0\ ; & \text{elsewhere}\end{cases}$$

Solution: Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from uniform population $U(\alpha, \beta)$. Therefore, likelihood function can be obtained as

$$L = f(x_1,\alpha,\beta).f(x_2,\alpha,\beta) \ldots f(x_n,\alpha,\beta)$$

$$= \underbrace{\left(\frac{1}{\beta-\alpha}\right)\left(\frac{1}{\beta-\alpha}\right)\ldots\left(\frac{1}{\beta-\alpha}\right)}_{n-\text{times}}$$

$$= \left(\frac{1}{\beta-\alpha}\right)^n$$

$$\Rightarrow \log L = -n\log(\beta-\alpha) \qquad \ldots (8)$$

Differentiating equation (8) partially with respect to α and β respectively, we get likelihood equations for α and β as

$$\frac{\partial}{\partial\alpha}(\log L) = 0 \Rightarrow \frac{n}{\beta-\alpha} = 0$$

and

$$\frac{\partial}{\partial\beta}(\log L) = 0 \Rightarrow \frac{-n}{\beta-\alpha} = 0$$

Both the equations give an inadmissible solution for α & β as β = α = ∞. So the method of differentiation fails. Thus, we have to use another method to obtain the desired result.

In such situations, we use basic principal of maximum likelihood, that is, we choose the value of the parameters α & β which maximize the likelihood function. If $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ is an ascending ordered arrangement of the observed sample, then $\alpha \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \le \beta$. Also, it can be seen $\beta \ge x_{(n)}$ and $\alpha \le x_{(1)}$. $\beta \ge x_{(n)}$ means, β takes values greater than or equal to $x_{(n)}$ and least value of β is $x_{(n)}$. Similarly, $\alpha \le x_{(1)}$ means, α takes values less than or equal to $x_{(1)}$ and maximum value of α is $x_{(1)}$. Now, likelihood function will be maximum when α is maximum and β is minimum. Thus, the minimum possible value of β consistent with the sample is $x_{(n)}$ and the maximum possible value of α consistent with the sample is $x_{(1)}$. Hence, L is maximum if $\beta = x_{(n)}$ and $\alpha = x_{(1)}$.

Thus, ML estimates for α and β are given by

$$\hat{\alpha} = x_{(1)} = \text{ Smallest sample observation}$$

and

$$\hat{\beta} = x_{(n)} = \text{ Largest sample observation}$$

Hence, ML estimators for α and β are $X_{(1)}$ and $X_{(n)}$ respectively.

The order statistics of a random sample $X_1, X_2, \ldots, X_n$ are the sample values placed in ascending order of magnitude. These are denoted by $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$

6.3.1 Properties of Maximum Likelihood Estimators

The following are the properties of maximum likelihood estimators:

1. A ML estimator is not necessarily unique.
2. A ML estimator is not necessarily unbiased.
3. A ML estimator may not be consistent in rare case.
4. If a sufficient statistic exists, it is a function of the ML estimators.

5. If T = t(X_1, X_2, …,X_n) is a ML estimator of θ and γ (θ) is a one to one function of θ, then γ(T) is a ML estimator of γ(θ). This is known as invariance property of ML estimator.
6. When ML estimator exists, then it is most efficient in the group of such estimators.

It is now time for you to try the following exercises to make sure that you get the concept of ML estimators.

E3) Prove that for the binomial population with density function

$$P\left[X = x\right] = {}^{n}C_{x}p^{x}q^{n-x}; \quad x = 1, 2, ..., n, q = 1 - p$$

the maximum likelihood estimator for p is $\bar{X}/n$.

E4) Obtain the ML estimate of θ for the following distribution

$$f\left(x, \theta\right) = \frac{1}{\theta}; \qquad 0 \le x \le \theta, \ \theta > 0$$

If the sample values are1.5, 1.0, 0.7, 2.2, 1.3 and 1.2.

E5) List any five properties of maximum likelihood estimators.

6.4 METHOD OF MOMENTS

The method of moments is the oldest but simple method for determining the point estimate of the unknown parameters. It was discovered by Karl Pearson in 1894. The application of this method was invariable continued until Prof. Ronald A. Fisher introduced maximum likelihood estimation method. The principle of this method consists of equating the sample moments to the corresponding moments of the population, which are the function of unknown population parameter(s). We equate as many sample moments as there are unknown parameters and solve these simultaneous equations for estimating unknown parameter(s). This method of obtaining the estimate(s) of unknown parameter(s) is called **"Method of Moments"**.

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a population whose probability density (mass) function is f(x, θ) with k unknown parameters, say, $\theta_1, \theta_2, \ldots, \theta_k$. Then the r^{th} sample moment about origin is

$$M'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

and about mean is

$$M_r = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r$$

While the r^{th} population moment about origin is

$$\mu'_r = E\left(X\right)^r$$

and about mean is

$$\mu_r = E\left(X - \mu\right)^r$$

Generally, the first moment about origin (zero) and rest central moments (about mean) are equated to the corresponding sample moments. Thus, the equations are

$$\mu_1' = M_1'$$

$$\mu_r = M_r \qquad ; r = 2, 3, \ldots, k$$

By solving these k equations for unknown parameters, we get the moment estimators.

Let us explain the concept of method of moments with the help of some examples.

Example 4: Find the estimator for λ by the method of moments for the exponential distribution whose probability density function is given by

$$f(x,\lambda) = \frac{1}{\lambda}e^{-x/\lambda}; \qquad x > 0,\ \lambda > 0$$

Solution: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from exponential distribution whose probability density function is given by

$$f(x,\lambda) = \frac{1}{\lambda}e^{-x/\lambda}; \qquad x > 0,\ \lambda > 0$$

We know that the first moment about origin, that is, population mean of exponential distribution with parameter λ is

$$\mu_1' = \lambda$$

and the corresponding sample moment about origin is

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Therefore, by the method of moments, we equate population moment with corresponding sample moment. Thus,

$$\mu_1' = M_1'$$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Hence, moment estimator for λ is $\bar{X}$.

Example 5: If $X_1, X_2, \ldots, X_m$ is a random sample taken from binomial distribution (n, p) where, n and p are unknown, obtain moment estimators for both n and p.

Solution: We know that the mean and variance of binomial distribution (n, p) are given by

$$\mu_1' = np \text{ and } \mu_2 = npq$$

Also the corresponding first sample moment about origin & second sample moment about mean (central moment) are

$$M_1' = \frac{1}{m}\sum_{i=1}^{m} X_i = \bar{X} \text{ and } M_2 = \frac{1}{m}\sum_{i=1}^{m}\left(X_i - \bar{X}\right)^2$$

Therefore, by the method of moments, we equate population moments with corresponding sample moments. Thus,

$$\mu_1' = M_1'$$

$$np = \frac{1}{m}\sum_{i=1}^{m} X_i = \bar{X} \qquad \ldots (9)$$

and

$$\mu_2 = M_2$$

$$npq = \frac{1}{m}\sum_{i=1}^{m}\left(X_i - \bar{X}\right)^2 = S'^2 \qquad \ldots (10)$$

We solve above equations (9) & (10) for n and p by dividing equation (10) by equation (9), we get

$$\hat{q} = \frac{S'^2}{\bar{X}}$$

Since p = 1 - q, therefore the estimator of p is

$$\hat{p} = 1 - \hat{q} = 1 - \frac{S'^2}{\bar{X}} = \frac{\bar{X} - S'^2}{\bar{X}}$$

Put the value of p in equation (9), we get

$$n\left(\frac{\bar{X} - S'^2}{\bar{X}}\right) = \bar{X}$$

$$\hat{n} = \frac{\bar{X}^2}{\bar{X} - S'^2}$$

Hence, moment estimators for p and n are $\frac{\bar{X} - S'^2}{\bar{X}}$ and $\frac{\bar{X}^2}{\bar{X} - S'^2}$ respectively

where,

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i \text{ and } S'^2 = \frac{1}{m}\sum_{i=1}^{m}\left(X_i - \bar{X}\right)^2.$$

Example 6: Show that moment estimator and maximum likelihood estimator are same of the parameter θ of the geometric distribution G(θ) whose pmf is

$$P\left[X = x\right] = \theta\left(1-\theta\right)^x; \qquad \theta > 0,\ x = 0, 1, 2, \ldots$$

Solution: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from G(θ) whose probability mass function is given by

$$P\left[X = x\right] = \theta\left(1-\theta\right)^x; \qquad x = 0, 1, 2, \ldots$$

Here, we first find the maximum likelihood estimator of θ. The likelihood function for parameter θ can be obtained as

$$L(\theta) = P\left[X = x_1\right].P\left[X = x_2\right] \ldots P\left[X = x_n\right]$$

$$= \theta(1-\theta)^{x_1}.\theta(1-\theta)^{x_2} \ldots \theta(1-\theta)^{x_n}$$

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

Taking log on both sides, we get

$$\log L = n\log\theta + \sum_{i=1}^{n} x_i \log(1-\theta) \qquad \ldots (11)$$

Differentiating equation (11) partially with respect to θ and equating to zero, we get

$$\frac{\partial}{\partial\theta}(\log L) = \frac{n}{\theta} + \sum_{i=1}^{n} x_i . \frac{1}{1-\theta}(-1) = 0$$

$$\Rightarrow \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i}{1-\theta} = 0 \Rightarrow \frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{1-\theta}$$

$$\Rightarrow n(1-\theta) = \theta\sum_{i=1}^{n} x_i \Rightarrow n - n\theta - \theta\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \theta\left(\sum_{i=1}^{n} x_i + n\right) = n$$

$$\Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i + n} = \frac{1}{\bar{x}+1}$$

Also, it can be seen that the second derivative, i.e.

$$\left.\frac{\partial^2}{\partial\theta^2}(\log L)\right|_{\theta=\frac{1}{\bar{x}+1}} < 0$$

Therefore, the ML estimator of θ is $\frac{1}{\bar{X}+1}$.

Now, we find the moment estimator of θ.

We know that the first moment about origin, that is, mean of geometric distribution is

$$\mu_1' = \frac{1-\theta}{\theta}$$

and the corresponding sample moment is

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Therefore, by the method of moments, we have

$$\mu_1' = M_1'$$

$$\Rightarrow \frac{1-\theta}{\theta} = \bar{X}$$

$$\Rightarrow \theta\bar{X} = 1-\theta \Rightarrow \theta\bar{X} + \theta = 1$$

$$\Rightarrow \theta(\bar{X}+1) = 1 \Rightarrow \hat{\theta} = \frac{1}{(\bar{X}+1)}$$

Thus, moment estimator of θ is $\frac{1}{\bar{X}+1}$.

Hence, the maximum likelihood estimator and moment estimator both are same in case of geometric distribution.

6.4.1 Properties of Moment Estimators

The following are the properties of moment estimators:

1. The moment estimators can be obtained easily.
2. The moment estimators are not necessarily unbiased.
3. The moment estimators are consistent because by the law of large numbers a sample moment (raw or central) is a consistent estimator for the corresponding population moment.
4. The moment estimators are generally less efficient than maximum likelihood estimators.
5. The moment estimators are asymptotically normally distributed.
6. The moment estimators may not be function of sufficient statistics.
7. The moment estimators are not unique.

6.4.2 Drawbacks of Moment Estimators

The following are the drawbacks of moment estimators:

1. This method is based on equating population moments with sample moments. But in some situations, like as cauchy distribution, the population moment does not exist therefore in such situations this method cannot be used.
2. This method does not, in general, give estimators with all the desirable properties of a good estimator.
3. The property of efficiency is not possessed by these estimators.
4. The moment estimators are not unbiased in general.
5. Generally, the moment estimators and the maximum likelihood estimators are identical. But if they do differ, then ML estimates are usually preferred.

Now, try to solve the following exercises to ensure that you have understood method of moments properly.

E6) Obtain the estimator of parameter λ when sample is taken from a Poisson population by the method of moments.

E7) Obtain the moment estimators of the parameters μ and σ^2 when the sample is drawn from normal population.

E8) Describe the properties and drawbacks of moment estimators.

6.5 METHOD OF LEAST SQUARES

The idea of least squares estimation emerges from the concept of method of maximum likelihood. Consider the maximum likelihood of the parameter μ when σ^2 is known on the basis of a random sample $Y_1, Y_2, \ldots, Y_n$ of size n taken from a normal population (μ, σ^2). The density function of normal population is given by

$$f\left(y,\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}; -\infty < y < \infty, -\infty < \mu < \infty\ , \sigma > 0$$

Then likelihood function for μ and σ^2 is

$$L\left(\mu,\sigma^2\right) = L = f\left(y_1,\mu,\sigma^2\right).f\left(y_2,\mu,\sigma^2\right)...f\left(y_n,\mu,\sigma^2\right)$$

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

Taking log on both sides, we have

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

By principle of maximum likelihood estimation, we have to maximize log L with respect to μ, and log L is maximum when $\sum_{i=1}^{n}(y_i-\mu)^2$ is minimum, i.e. sum of squares $\sum_{i=1}^{n}(y_i-\mu)^2$ must be least.

The method of least squares is mostly used to estimate the parameter of linear function. Now, suppose that the population mean μ is itself a linear function of parameters $\theta_1, \theta_2, \ldots, \theta_k$, that is,

$$\mu = x_1\theta_1 + x_2\theta_2 + ... + x_k\theta_k$$

$$= \sum_{j=1}^{k} x_i\theta_i$$

where, x_i's are not random variables but known constant coefficients of unknown parameter θ_i's for forming a linear function of θ_i's.

We have to minimize

$$E = \sum_{i=1}^{k}\left(y_i - \sum_{i=1}^{k} x_i\theta_i\right)^2 \text{ with respect to } \theta_i.$$

Hence, method of least squares gets its name from the minimization of a sum of squares. The principle of least squares states that we choose the values of unknown population parameters $\theta_1, \theta_2, \ldots, \theta_k$, say, $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ on the basis of observed sample observations $y_1, y_2, \ldots, y_n$ which minimize the sum of squares of deviations $\sum_{i=1}^{k}\left(y_i - \sum_{i=1}^{k} x_i\theta_i\right)^2$.

Note: The method of least squares has already been discussed in Unit 5 of MSL-002 and further application of this method in estimating the parameters of regression models is discussed in specialisation courses.

6.5.1 Properties of Least Squares Estimators

Least squares estimators are not so popular. They possess some properties which are as follows:

1. Least squares estimators are unbiased in case of linear models.
2. Least squares estimators are minimum variance unbiased estimators (MVUE) in case of linear models.

Now, try the one exercise.

E9) Describe the two properties of least squares estimators.

We now end this unit by giving a summary of what we have covered in it.

6.6 SUMMARY

After studying this unit, you must have learnt about:

1. The point estimation.
2. Different methods of finding point estimators.
3. The method of maximum likelihood and its properties.
4. The method of moments, its properties and drawbacks.
5. The method of least squares and its properties.

6.7 SOLUTIONS /ANSWERS

E1)

(i) Here, investigator estimates the average income Rs. 1.5 lack i.e. he / she estimate average income with a single value, therefore, the investigator used point estimation technique.

(ii) The product manager estimates the average life of electric bulbs with the help of two values 800 hours and 1000 hours, therefore, the product manager used interval estimation technique.

(iii) A pathologist estimates the mean time required completing a certain analysis with the help of two values 30 minutes and 45 minutes, therefore, he/she used interval estimation technique.

E2) Refer Sub-section 6.2.1.

E3) Let $X_1, X_2, ..., X_m$ be a random sample of size m taken from B(n, p) whose pmf is given by

$$P[X = x] = {}^nC_x p^x q^{n-x}; \quad x = 0,1,..., n \;\&\; q = 1-p$$

The likelihood function for p can be obtained as

$$L(p) = L = P[X = x_1].\ P[X = x_2] \ldots P[X = x_m]$$

$$= {}^nC_{x_1} p^{x_1} q^{n-x_1}.\,{}^nC_{x_2} p^{x_2} q^{n-x_2} \ldots {}^nC_{x_m} p^{x_m} q^{n-x_m}$$

$$= \prod_{i=1}^{m} \left({}^nC_{x_i}\right) p^{\sum_{i=1}^{m} x_i} q^{\sum_{i=1}^{m}(n-x_i)}$$

Taking log on both sides, we have

$$\log L = \log \prod_{i=1}^{m} \left({}^nC_{x_i}\right) + \sum_{i=1}^{m} x_i \log p + \sum_{i=1}^{m} (n - x_i) \log(1-p) \; [\because q = 1-p]$$

Differentiate partially with respect to p and equating to zero, we have

$$\frac{\partial}{\partial p}(\log L) = 0 + \sum_{i=1}^{m} x_i \left(\frac{1}{p}\right) + \sum_{i=1}^{m} (n - x_i)\left(\frac{1}{1-p}\right)(-1) = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{m} x_i}{p} = \frac{\sum_{i=1}^{m} (n - x_i)}{1-p}$$

$$\Rightarrow \frac{\sum_{i=1}^{m} x_i}{p} = \frac{nm - \sum_{i=1}^{m} x_i}{1-p}$$

$$\Rightarrow \sum_{i=1}^{m} x_i - p\sum_{i=1}^{m} x_i = nmp - p\sum_{i=1}^{m} x_i$$

$$\Rightarrow \hat{p} = \frac{\sum_{i=1}^{m} x_i}{nm} = \frac{\overline{x}}{n} \qquad \left[\because \overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i\right]$$

Also, it can be seen that the second derivative, i.e.

$$\frac{\partial^2}{\partial p^2}(\log L)\bigg|_{p=\frac{\overline{x}}{n}} < 0$$

Hence, ML estimator for parameter p is $\overline{X}/n$.

E4) Let $X_1, X_2, \ldots, X_n$ be a random sample taken from given $U(0, \theta)$ distribution. The likelihood function for θ is given by

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

$$= \left(\frac{1}{\theta}\right)\left(\frac{1}{\theta}\right)\ldots\left(\frac{1}{\theta}\right) = \left(\frac{1}{\theta}\right)^n$$

Taking log on both sides, we have

$$\log L = -n\log\theta \qquad \ldots (12)$$

Differentiate equation (12) partially with respect to θ, we get

$$\frac{\partial}{\partial\theta}(\log L) = \frac{-n}{\theta} = 0, \text{ has no solution for } \theta.$$

So ML estimate cannot be found by differentiation. Therefore, by the principle of ML estimation, we choose the value of θ which maximize likelihood function. Hence, we choose θ as small as possible.

If $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, is an ordered sample from this population, then $0 \le x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \le \theta$. Also, it can be seen that $\theta \ge x_{(n)}$ which means that θ takes values greater than or equal to $x_{(n)}$ and minimum value of θ is $x_{(n)}$. Now, likelihood function will be maximum when θ is minimum. The minimum value of θ is $x_{(n)}$. Therefore, ML estimate of θ is the maximum observation of the sample, that is, $\hat{\theta} = x_{(n)}$

Here, the given random sample is 1.5, 1.0, 0.7, 2.2, 1.3 and 1.2. Therefore, the ordered sample is 0.7 < 1.0< 1.1 < 1.3 < 1.5 < 2.2. Here, the maximum observation of this sample is 2.2, therefore maximum likelihood estimate of θ is 2.2.

E5) Refer Section 6.3.1.

E6) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from Poisson population whose probability mass function is given by

$$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}; \qquad x = 0, 1, 2, ... \& \lambda > 0$$

We know that for Poisson distribution

$$\mu_1' = \lambda$$

and the corresponding sample moment is

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Therefore, by the method of moments, we equate population moment with corresponding sample moment. Thus,

$$\mu_1' = M_1'$$

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Hence, moment estimator for λ is $\bar{X}$.

E7) Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from normal population $N(\mu, \sigma^2)$, whose probability density function is given by

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}; -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

We know that for $N(\mu, \sigma^2)$

$$\mu_1' = \mu \quad \text{and} \quad \mu_2 = \sigma^2$$

and the corresponding sample moments are

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \text{ and } M_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

Therefore by the method of moments, we equate population moments with corresponding sample moments. Thus,

$$\mu_1' = M_1'$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

and

$$\mu_2 = M_2$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = S'^2$$

Hence, moment estimators for μ and σ^2 are $\bar{X}$ and S'^2 respectively.

E8) Refer Sub-sections 6.4.1and 6.4.2.

E9) Refer as Section 6.5.1.