UNIT 7 POINT ESTIMATION

Structure

7.1 INTRODUCTION

So far previous blocks you have studied random variables and their probability distribution from theoretical point of view. You have learnt to compute the probabilities of various events associated with random variables and means, variances, moments, correlations coefficients etc. of random variables from their probability distributions. In practical situations the probability distribution of a random variable of interest may seldom be known. Even when the form of probability density function or cummulative distribution function of random variable is known, it may involve unknown parameters such as μ, σ^2 in normal distribution, or p in geometric distribution and so any of above quantities computed from these probability distributions become unknown. Thus for a random variable of interest you need to estimate the values of unknown parameters present in its probability distribution or the probability distribution itself if unknown, with the help of repeated observations on the random variables. For this purpose you will require the application of methods of statistical inference. When the problem is of obtaining single estimate of unknown parameters, such as μ, σ^2, p etc, or the probability density function itself, using observed data on the random variable, the statistical methods used for this purpose are dealt under the theory of point estimation. You will be introduced to this theory in this unit. Before we discuss the problem of point estimation, you will at the outset get conversent with basic statistical terminology and concepts needed in this unit and the remaining units in this block.

Objectives

After reading this unit, you will be able to

* recognise a problem needing statistical investigation,
* use standard statistical terminology,
* derive point estimators for unknown population parameters,
* construct moment estimator,

* construct maximum likelihood estimator,

* apply criterion of unbiasedness, consistency and minimum variance to judge a good estimator, and

* derive asymptotic probability distribution of maximum likelihood estimator.

7.2 POPULATION AND SAMPLE

All problems that call for a statistical solution have certain common characteristics. Let us identify these by considering a typical problem.

It is well known that hardness of brick is an important characteristic in building construction work. The hardness X of a brick manufactured at a plant is a random variable and therefore you may like to have some idea about the mean hardness $E(X)$ in order to judge the quality of bricks. Now if the manufacturing plant is new we shall have no knowledge about the probability density function (*pdf*) of X and we must discover it. But if it is known from experience from other plants that X has usually a normal distribution $N(\mu, \sigma^2)$, then $E(X) = \mu$, and μ is the parameter of interest which is unknown . Clearly the totality of bricks manufactured or to be manufactured at the plant is infinitely large and it is not feasible to measure the hardness X of each brick. Furthermore, if the hardness test is destructive then in the process of measuring hardness all bricks will be destroyed. Thus we shall only take a representative sample of manufactured bricks from the plant, measure X for each brick in the sample and utilise this observed data to estimate *pdf* of X or the parameter μ as the case may be. This is inherently a statistical problem. From this can you now identify the characteristics common to the problem needing statistical solutions ?

These are as follows :-

* There is a large collection of objects under investigation we shall call it a **population.**

  In the above problem the population under study is the collection of bricks manufactured or to be manufactured at the plant.

  We wish to study behaviour of a random characteristic, say X associated with the objects in the population. We may be interested in probability distribution of X, or m numerical characteristics such as mean $E(X)$ or variance $V(X)$ of random variable X.

  Let us denote by $f(x; \theta)$ the probability density function (*pdf*) or probability mass function (*pmf*) of random variable X depending on whether X is continuous or discrete. We call the probability distribution $f(x; \theta)$ of X as the **population distribution**. The constant θ, which may be real or a vector, present in population distribution is called **population parameter**.

  In the illustrative problem if hardness X is assumed to have normal $N(\mu, \sigma^2)$ distribution than population distribution is $N(\mu, \sigma^2)$. Mean hardness μ is the population parameter of interest though population distribution is dependent on vector parameter $\theta = (\mu, \sigma^2)$.

* The population is too large to study in entirty. We must draw inferences about the population, infact about the population distribution or population parameters by selecting a representative sample of say n objects selected at random from the population and measuring X for each object in the sample.

Suppose $X_1, X_2, \ldots, X_n$ are the values of X for n objects in the sample. Prior to actual selection of objects, value X_i of X for i th object to be included in the sample is a random variable which is a copy of X i.e. X_i has the same probability distribution as X. Thus ($X_1, X_2, \ldots, X_n$) is a set of independently and identically distributed random variables each having the same probability distribution $f(x ; \theta)$ as that of population. The set ($X_1, \ldots, X_n$) is called a random sample of size n on X, or from probability distribution $f(x ; \theta)$.

Continuiting with illustrative problem of hardness of bricks you may note that if X_i is the hardness of ith brick in the sample, then ($X_1, X_2, \ldots, X_n$) is random sample of size n on hardness X, or from probability distribution $N(\mu, \sigma^2)$.

* When we have actually selected the sample of objects measure X for these, we get a particular value of X_i ($i = 1, 2, \ldots, n$) which we denote by X_i. Then ($x_1, x_2, \ldots, x_n$) is a particular value taken by random sample ($X_1, X_2, \ldots, X_n$) and is called observed sample (or observed data) of size n. This is the data which is used in statistical inference on the population under study.

In the illustrative problem let 5 bricks chosen at random from the production line yield hardness values as (45.5, 58.1, 40.1, 70.4, 60.2) measured on a continuous scale between 0 and 100. Then this is an observed sample of size 5 and is regarded as particular value of random sample (X_1, X_2, X_3, X_4, X_5) of size 5 from normal $N(\mu, \sigma^2)$ population.

7.2.1 Sample Statistics

The observed sample ($x_1, x_2, \ldots, x_n$) contains information about the population distribution or its parameters. The desired information can be condensed or summarized into a single quantity by choosing a sensible function $H(x_1, \ldots, x_n)$ of observed sample ($x_1, x_2, \ldots, x_n$). $H(x_1, \ldots, x_n)$ is a particular value taken by random variable $H(X_1, \ldots, X_n)$.

Definition:

A function $H(X_1, \ldots, X_n)$ of random sample ($X_1, \ldots, X_n$) is called a sample statistic or simply a statistic.

A statistic H being a random variable, has a probability distribution. This probability distribution is called sampling distribution of statistic H. You have already been introduced to some sampling distributions in the previous Block II.

* Sample Mean : $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
* Sample Variance : $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
* Maximum of Sample : $M = \max(X_1, \ldots, X_n)$
* Minimum of Sample : $m = \min(X_1, \ldots, X_n)$
* Sample Range : $R = M - m$
* Sample Moments : $m_r = \frac{1}{n} \sum X_i^r$

  $r = 1, 2, \ldots$

These sample statistics provide information about population characteristics such as population mean μ, population variance σ^2, population skewness etc. Since the sample statistics are quantities based on observations on a small portion of objects constituting the population under study. For instance, the population mean $\mu = E(X)$ is the average value of X taken over the entire population, whereas sample mean $\bar{X}$ is average of X over a sample of n objects chosen from the population. So you will not be wrong if you hope that $\bar{X}$ is close to μ. Similarly, sample variance and sample range contain information about the variability in X in the population which is measured by $V(X) = \sigma^2$

Study the following examples now for gaining more clarity on what you have learnt in this section.

Example 1 :

In order to decide on the number of remote terminals needed in a computer system, it is necessary to make a study of the length of time spent by users at terminals per session. The length of time spent by a user is a random variable having an exponential distribution with parameter α (say).

(i) What is the population under study ?

(ii) What random characteristic of objects in population is of interest ?

(iii) What is the population distribution ?

(iv) What is the parameter of interest ?

(v) What is a random sample of size 5 from the population ? Give a statistic based on this random sample that is useful for the parameter of interest.

Solution :

(i) Population under study is collection of all present and future users of computer terminals.

(ii) X, the time spent per session on computer terminals by a user is the random characteristic of interest.

(iii) Population distribution is exponential distribution with parameter α The *pdf* of population distribution is $f(x;\alpha) = \alpha e^{-\alpha x}, x > 0$.

(iv) The mean time per session spent on computer terminals by a user is $E(X) = \frac{1}{\alpha}$, the knowledge of which is important in the decision on the number of terminals to be provided. Thus $E(X)$ or equivalently α is the parameter of interest in the problem.

(v) A set X_1, X_2, X_3, X_4, X_5 of independently and identically distributed random variables each having the same exponential distribution with parameter α is the random sample of size 5 on X. For parameter $E(X)$, the sample mean $\bar{X} = \frac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5)$ is a useful sample statistics, so for α it may be $\frac{1}{\bar{X}}$.

Example 2 :

Performance of reflective highway signs depends on the proper allignment of headlights of the automobiles in operation. It may be of interest to study the proportion of automobiles in operation in a city that have proper alignment of head lights.

(i) What is the population under study ?

(ii) What random characteristic of objects in the population is of interest ?

(iii) What is the population distribution ? Give the population parameter of interest ?

(iv) Can you obtain an observed sample of size 10 ? If yes, how ?

(v) Define a statistics that is useful for the parameter of interest and compute it particular from an observed sample obtained as in part (iv).

Solution :

(i) Population under study is the collection of all automobiles in operation in the city.

(ii) The random characteristic X of an object that is an automobile here can be defined by

$$X = \begin{cases} 1, & \text{if automobile has proper allignment} \\ 0, & \text{if automobile does not have proper allignment} \end{cases}$$

(iii) Let p be the proportion of automobiles in the city that have proper allignment. Then X has Bernoulli's distribution

$$P(X = 0) = 1-p, p(X = 1) = p.$$

Which is the population distribution and the population parameter of interest is p, which is also the population mean.

(iv) To obtain an observed sample on X, select 10 automobiles at random from the population of automobiles in operation in the city. Check the alignment of 10 automobiles chosen. Record a1, if allignment is o.k. and record a 0, otherwise for each of 10 automobiles. A typical observed sample is a sequence of 0's or 1's of length 10, say, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0.

(v) Sample mean

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\#\text{ of } 1's}{n} = \hat{p} \text{ (say)}$$

is the proportion of automobiles having proper allignment of headlights in the sample of size n. Since population parameter of interest is population mean p, the sample statistic useful for it is sample mean $\overline{X}$, or sample proportion $\hat{p}$. A particular value of $\hat{p}$ for the observed sample given in (iv) is

$$\hat{p} = \frac{6}{10}$$

Example 3 :

When running computer programmes on time sharing basis, cost varies from session to session. Let X be the cost in rupees per session to a user. A observed sample of size on X is

$$x_1 = 3900, x_2 = 4000, x_3 = 4060, x_4 = 4100, x_5 = 4500, x_6 = 3950,$$

$$x_7 = 4010, x_8 = 4080, x_9 = 4150$$

Compute values of sample mean $\overline{X}$, sample variance s^2, sample range R and sample 3rd moment m_3 for this sample.

Solution :

Here

$$n = 9, \sum x_i = 36750, \sum x_i^2 = 1.50305 \times 10^8$$

$$\sum x_i^3 = 6.1579 \times 10^{11}.$$

Thus

$$\overline{x} = \frac{36750}{9} = 4083.3 \text{ Rs.}$$

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

$$= \frac{1}{8} \left[1.50305 \times 10^8 - \frac{(36750)^2}{9} \right]$$

$$= 3.03125 \times 10^4 \text{ (Rs.)}^2$$

$$R = \max x_i - \min x_i = 4500 - 3900$$

$$= 600 \text{ Rs.}$$

$$m_3 = \frac{6.1579 \times 10^{11}}{9} = 6.84211 \times 10^{10} \text{ (Rs.)}^3.$$

Check Your Progress A

E1

A factory has a large number of workers. The management of factory is planning to replace an old pension scheme for the workers by new scheme. The management will like to know the proportion of workers that are in favour of new scheme before taking any final decision.

(i) The population under study is ....

(ii) The random variable X of interest in the population is defined by ......
....

(iii) The parameter of interest in the population is ....

(iv) The population distribution is ....

(v) A sample statistic that provides information about parameter of interest is
....

A high yielding variety of rice seeds has been developed for sowing in a region. An agricultural scientist wishes to have some idea about expected yield per acre obtained from these seeds. Yield per acre can be assumed to have a normal distribution with mean μ and variance σ^2. Ten farmers were chosen at random and seeds distributed to them. They reported yield/acre as below

3502	3215	2826	4003	3502
4500	3885	3628	4223	3495

(i) Population under study is . . . .

(ii) Population distribution is . . . .

(iii) Population parameter of interest is . . . .

(iv) A statistic useful for the parameter of interest is . . . .

(v) The value of statistic in (iv) for the observed data given above is . . . .

(vi) The value of sample variance for the data is . . . .

7.3 POINT ESTIMATION

You now know that a random sample $X_1, X_2, \ldots, X_n$ from a population distribution $f(x, \theta)$ can be used to define sample statistics which contain information about the population distribution or population parameter θ. If θ is unknown you may like to use the information contained a statistic to provide an approximate value or an estimate for θ. In point estimation we learn how to do this.

7.3.1 Point Estimator

To find an estimate of unknown population parameter θ. We adopt the following procedure :

* Choose an appropriate sample statistic $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ based on random sample $(X_1, X_2, \ldots, X_n)$ from the population.

* Obtain an observed sample $(x_1, x_2, \ldots, x_n)$ which you know is a particular value of random sample $(X_1, X_2, \ldots, X_n)$.

* Compute $\hat{\theta}(x_1, x_2, \ldots, x_n)$, the value assumed by chosen sample statistic $\hat{\theta}$ for the observed sample $(x_1, x_2, \ldots, x_n)$. Then value $\hat{\theta}(x_1, x_2, \ldots, x_n)$, is taken as the estimate of θ based on observed sample $(x_1, x_2, \ldots, x_n)$.

Definition :

A sample statistics $\hat{\theta}(X_1, X_2, ..., X_n)$ used for estimating θ is called a point estimator of unknown population parameter θ. A particular value $\hat{\theta}(x_1, x_2, ..., x_n)$ assumed by point estimator $\hat{\theta}$ for an observed sample $(x_1, x_2, ..., x_n)$ is called point estimate of θ.

Note :

A point estimator is simply referred as an estimator. An estimator is a random variable since it is a function of random sample, while an estimate is a particular value of this random variable. Thus an estimator gives infinity many estimates of the same parameter θ depending on observed samples.

As an illustration, suppose you are asked to estimate the mean hardness μ of bricks manufactured by a plant, where hardness X is assumed to have a normal distribution $N(\mu, \sigma^2)$. Normal distribution being symmetrical both its mean as well as its median are equal to μ. Since population mean and median each equals the parameter μ, you may be tempted to consider two sample statistics viz sample mean $\overline{X}$ and sample median Md based on random sample$(X_1, X_2, \ldots, X_n)$ from $N(\mu, \sigma^2)$ as point estimators for μ. Sample median Md is defined as follows. Arrange the random sample in non decreasing order as

$$X_{(1)} \leq X_{(2)} \leq X_{(n)}$$

that is, $X_{(i)}$ is ith largest among $X_1, X_2, ..., X_n$ and is called ith order statistic.

Then

$$Md = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n+1}{2}\right)}\right) & \text{if } n \text{ is even} \end{cases}$$

Thus you notice that

$$\hat{\mu}_1 = \overline{X} \text{ and } \hat{\mu}_2 = M_d$$

can be two point estimators of populatlion parameter μ. In fact, you can think of many more sample statistics that you may intuitiuely feel as appropriate for being considered as point estimator of μ. Just try it.

Now suppose you actually measure the hardness of 5 bricks chosen at random from the plant as 45.5, 58.1, 40.1, 70.4, 60.2.

Based on this observed sample the estimate of μ if $\hat{\mu}_1$ is chosen as estimator is 54.86, whereas if we choose $\hat{\mu}_2$ as estimator it is 58.1, the third highest observation in the sample.

Since one can construct several point estimators for the same population parameter θ, some natural questions that will arise in your mind may be

What are the properties of a good estimator ?

Which point estimator is relatively better among two ?

How can one construct estimators with good properties ?

You will get answers to these question gradually as you read the rest of this unit. Let us start with some desirable properties for point estimators.

Obviously you would like a point estimator $\hat{\theta}$ of parameter θ to generate estimates lying around the true unknown value of θ with their average coinciding with this value of θ. Furthermore the estimates generated should be close to true value of parameter θ whatever the true value of θ. Since $\hat{\theta}$ is a random variable these requirements will be fulfilled if $\hat{\theta}$ satisfies the following two conditions whatever value of parameter θ may be.

(a) $E(\hat{\theta}) = \theta$

(b) $V(\hat{\theta})$ is small.

These two requirements respectively give rise to concepts of unbiased estimator and minimum variance unbiased estimator.

7.3.2 Unbiased Estimator

Definition :

An estimator $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ is called unbiased for parameter θ if

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

for all θ. Otherwise, $\hat{\theta}$ is called biased and

$$B(\theta) = E(\hat{\theta}) - \theta$$

is called bias of $\hat{\theta}$ at the point θ.

Note :

Unbiasedness of an estimator $\hat{\theta}$ does not imply that estimates $\hat{\theta}(x_1, \ldots, x_n)$ generated by it are close to θ, but only that their average coincides with θ and some estimates generated by unbiased estimator $\hat{\theta}$ may be far away from θ.

Most often population mean μ and variance σ^2 are the two important quantities that you would like to estimate if unknown. We now prove that sample mean $\overline{X}$, and sample variance s^2 are unbiased estimators of μ and σ^2 respectively for any population. These results are contained in the following theorems.

Theorem 1:

Let population mean be μ and population variance be σ^2. Let $\overline{X}$ be the sample mean based on a random sample $(X_1, X_2, \ldots, X_n)$ from the population, then

$$E(\overline{X}) = \mu, \; V(\overline{X}) = \frac{\sigma^2}{n}$$

for any μ and σ^2.

Proof :

By definition of random sample $E(X_i) = \mu$ and $V(X_i) = \sigma^2$ for any i and X_i are independent.

Hence

$$E(X) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

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

$$= \frac{1}{n} . n\mu$$

$$= \mu$$

and

$$V(\overline{X}) = V\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V(X_i)$$

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

$$= \frac{\sigma^2}{n}$$

This completes the proof.

Theorem 2 :

Let population variance be σ^2. Let s^2 be the sample variance based on random sample $(X_1, X_2, \ldots, X_n)$ from the population then $E(s^2) = \sigma^2$, for any σ^2

Proof :

Let μ be the population mean. By definition of s^2

$$(n-1)s^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

$$= \sum_{i=1}^{n} [(X_i - \mu) - (\overline{X} - \mu)]^2$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X} - \mu) \sum_{i=1}^{n} (X_i - \mu) + n(\overline{X} - \mu)^2$$

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

Therefore,

$$E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\overline{X} - \mu)^2\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - n . \frac{\sigma^2}{n}\right]$$

$$= \sigma^2$$

We have used the result $V(\overline{X}) = E(\overline{X} - \mu)^2 = \frac{\sigma^2}{n}$ from the Theorem 1. The proof is completea.

It is possible that an unbiased estimator does not exist for a parameter. Also unbiased estimator may not be unique even when it exists. For example consider the statistics, the weighted sample mean

$$\overline{X}_w = \sum_{i=1}^{n} a_i X_i ,$$

where $\sum_{i=1}^{n} a_i = 1$ and a_i are constants.

Then

$$E(\overline{X}_w) = \sum_{i=1}^{n} a_i E(X_i) = \mu \sum_{i=1}^{n} a_i = \mu$$

Hence $\overline{X}_w$ is unbiased estimator of population mean m for all choices of a_i subject to restriction

$$\sum_{i=1}^{n} a_i = 1.$$

Sample mean X is just one unbiased estimator in this class which you get when $a_i = \frac{1}{n}$ for all i. Now the question that may concern you is how to make a plausible choice of an unbiased estimator if there are more than one unbiased estimators for the same parameter. For this purpose we turn our attention to the variance of unbiased estimators.

7.3.3 Minimum Variance Unbiased Estimator

It was indicated in discussion before unbiased estimator that in order to ensure that the estimates generated by an estimator $\hat{\theta}$ for parameter θ are close to θ you will need an estimator with small variance. Thus suppose you have two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ for a parameter θ i.e.

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$$

for all θ . Now if $V(\hat{\theta}_1) < V(\hat{\theta}_2)$ for all θ , then if would mean that estimates generated by estimator $E(\hat{\theta}_1)$ tend to be more closer to the true value of parameter θ whatever it may be, in comparison to the estimates generated by $\hat{\theta}_2$.

This fact is shown in the Figure 7.1 by sketching graphs of *pdf's* $f(\hat{\theta}_1)$ and $g(\hat{\theta}_2)$ of $\hat{\theta}_1$ and $\hat{\theta}_2$ clearly in this situation you will prefer $\hat{\theta}_1$ over $\hat{\theta}_2$ in estimating θ.

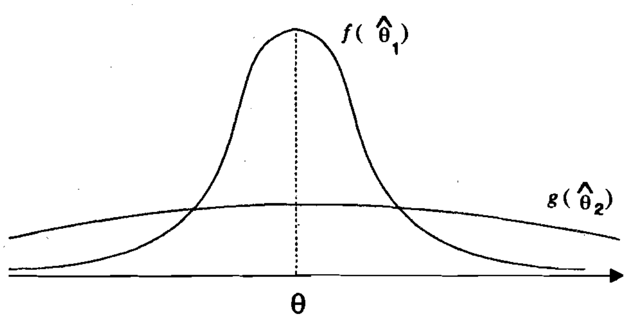

Figure 7.1

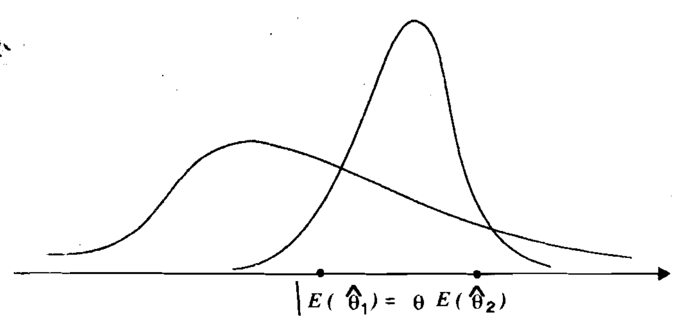

Figure 7.2

Consider another situation as sketched in Figure 7.2. Here $\hat{\theta}_1$ is unbiased estimator of θ but $\hat{\theta}_2$ has a bias

$$B(\theta) = E(\hat{\theta}_2) - \theta$$

but

$$V(\hat{\theta}_2) < V(\hat{\theta}_1).$$

Thus the choice between $\hat{\theta}_1$ and $\hat{\theta}_2$ is not clear. If the emphasis is on accuracy of estimate generated and a small amount of bias can be tolerated then you may prefer $\hat{\theta}_2$ over $\hat{\theta}_1$ though former is biased. For example, if you are to estimate mean yield of crop per acre a small bias in estimator can be ignored if its variance is very small but this may not be the case when you are to estimate sulpher content in ground water to be used for drinking. So choice between estimators in such situations is dictated by the purpose for which estimator is needed. However, if we have to choose between several unbiased estimators then the estimator with least variance, whatever be the true value of parameter θ, will be preferred. This leads to the following definition.

Definition :

An unbiased estimator $\hat{\theta}(X_1, ..., X_n)$ of population parameter θ is called minimum variance unbiased estimator of θ if

$$V(\hat{\theta}) \leq V(\hat{\theta}^*)$$

for all θ and every other unbiased estimator $\hat{\theta}^*$ of θ.

The minimum variance unbiased estimators may not exist. You will not learn the problem of existence of these or of uniqueness when they exist here. The following example will make use of above concepts.

Example 4 :

The time to failure T of a component has exponential distribution with $pdf f(t, \alpha) = \alpha e^{-\alpha t}, t > 0$, where α is unknown parameter. Let $T_1, T_2, ..., T_n$ be a random sample on T. To estimate the meantime to failure of the component $E(T)$, consider two estimators defined by

$$\hat{\mu}_1 = \overline{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$$

$$\hat{\mu}_2 = n\,T_{(1)}, \text{ where } T_{(1)} = \min(T_1, T_2, ..., T_n)$$

Show that both $\hat{\mu}_1$ and $\hat{\mu}_2$ are unbiased for $E(T)$. If you have choice between these two unbiased estimators which one will prefer.

Solution :

For exponential distribution the mean and variance are $E(T) = 1/\alpha$ and $V(T) = 1/\alpha^2$. From the result of Theorem 1, it follows that sample mean $\overline{T}$ or $\hat{\mu}_1$ is unbiased estimator of population mean $1/\alpha$ and

$$V(\hat{\mu}_1) = \frac{1}{n\alpha^2}$$

Now it remains to determine

$$E(\hat{\mu}_2)$$

and

$$V(\hat{\mu}_2).$$

Since

$$E(\hat{\mu}_2) = nE(M)$$

and

$$V(\hat{\mu}_2) = n^2 V(M)$$

We must determine $E(M)$ and $V(M)$. For this purpose we need *pdf* of M which you have learnt in Block II. Do you remember that ? If not, let us derive it again. The distribution function of M is given by

$$\begin{aligned} F(m) &= P(M \le m) \\ &= 1 - P(M > m) \\ &= 1 - P(T_1 > m, T_2 > m, \ldots, T_n > m) \\ &= 1 - \prod_{i=1}^{n} P(T_i > m) \\ &= 1 - (e^{-\alpha m})^n \\ &= 1 - e^{-n\alpha m}, m > 0 \end{aligned}$$

which is the distribution function of exponential distribution with parameter $n\alpha$ Thus

$$E(M) = \frac{1}{n\alpha} \text{ and } V(M) = \frac{1}{n^2\alpha^2}$$

Hence

$$E(\hat{\mu}_2) = n \cdot \frac{1}{n\alpha} = \frac{1}{\alpha}$$

and

$$V(\hat{\mu}_2) = n^2 \cdot \frac{1}{n^2\alpha^2} = 1/\alpha^2$$

Which shows $\hat{\mu}_2$ is also unbiased estimator of population mean $\frac{1}{\alpha}$. Comparing variances of $\hat{\mu}_1$ and $\hat{\mu}_2$, you find that for

$$n > 1, V(\hat{\mu}_1) < V(\hat{\mu}_2)$$

for all $\alpha > 0$. So we conclude that $\hat{\mu}_1$ is better unbiased estimator than $\hat{\mu}_2$ if only consideration is statistical.

Note :

Some times economic consideration may also dictate the choice of estimator. In Example 4, in order to obtain complete sampie you will have to put n components to failure test and wait till all components fail. Then T_i (i = 1, 2, ..., n) the time to failure of ith component is the ith observation in the sample. The estimator $\hat{\mu}_1$ can be computed only after complete random sample ($T_1, T_2, \ldots, T_n$) has been observed. Whereas, for computation of $\hat{\mu}_2$ one need not wait for that long as now the estimate can be obtained as soon as any one component out of n on test fails and operation on remaining ($n - 1$) components can be stopped. Therefore, if $\hat{\mu}_2$ is used as an estimator you will save both in time as well as money since now only one component is allowed to fail leaving remaining in working condition. Thus economic consideration may force you to prefer estimator $\hat{\mu}_2$ over $\hat{\mu}_1$.

7.3.4 Consistent Estimator

One may argue that as sample sige n increases, we tend to cover larger and larger portion of population and it is intuitively reasonable to expect that as n increases to ∞ any good estimator $\hat{\theta}$ of the parameter θ should ultimately converge to true value of θ in some sense. For example, this is the manner in which sample mean $\bar{X}$ behaves as an estimator for parameter μ, the population mean. From Theorem 1, we know $E(\bar{X}) = \mu$ and $V(\bar{X}) = \frac{\sigma^2}{n}$. So you see that $V(\bar{X}) \to 0$ as sample size $n \to \infty$, which means the variation of $\bar{X}$ around μ as $n \to \infty$ reduces to nil. In other words $\bar{X}$ tends to take constant value μ as $n \to \infty$. We say $\bar{X}$ is a consistent estimator of μ. Consistency is just another desirable property for judging an estimation. A formal definition consistent estimator is as follows.

Definition :

An estimator $\hat{\theta}(X_1, X_2, \ldots, X_n)$ based on a random smple of size n is called a consistent estimator for a parameter θ, if for all $\varepsilon > 0$.

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

The definition in simple words states that if sample size n is made very large then with high probability close to one, the absolute difference between a consistent estimator $\hat{\theta}$ and the parameter θ which it estimates will be less than ε, which can be chosen as small as we wish. Consequently, $\hat{\theta}$ is near to θ with probability one when n is infinitely large. We say $\hat{\theta}$ converges to θ with probability one.

In many practical problems the above definition is hard to apply for checking consistency of an estimator. The following theorem may be useful in some problems.

Theorem 3 :

An estimator $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is consistent for a parameter θ, if the following conditions hold

(1) $\lim_{n \to \infty} E(\hat{\theta}) = \theta$

(2) $\lim_{n \to \infty} V(\hat{\theta}) = 0$

Proof :

By Chebychev's inequality that you have learnt in Block II, for any $\varepsilon > 0$

$$P(|\hat{\theta}-\theta| \geq \varepsilon) \leq \frac{E(\hat{\theta}-\theta)^2}{\varepsilon^2}$$

Furthermore, since

$$\begin{aligned} E(\hat{\theta}-\theta)^2 &= E[(\hat{\theta}-E(\hat{\theta}))+(E(\hat{\theta})-\theta)]^2 \\ &= E(\hat{\theta}-E(\hat{\theta}))^2+2(E(\hat{\theta})-\theta)E(\hat{\theta}-E(\hat{\theta}))+E(\hat{\theta})-\theta)^2 \\ &= V(\hat{\theta})+0+(E_{(\hat{\theta})}-\theta)^2 \end{aligned}$$

We have by conditions given in the theorem that

$$\lim_{n \to \infty} E(\hat{\theta}-\theta)^2 = 0$$

Therefore

$$\lim_{n \to \infty} P(|\hat{\theta}-\theta| \geq \varepsilon) \leq \lim_{n \to \infty} \frac{E(\hat{\theta}-\theta)^2}{\varepsilon^2} = 0$$

Since probability is always ≥ 0, we must have

$$\lim_{n \to \infty} P(|\hat{\theta}-\theta| \geq \varepsilon) = 0 \text{ for any } \varepsilon > 0$$

The proof is complete.

Example 5 :

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population having a mean μ and variance σ^2.

Show that

$$\bar{X} = \frac{2}{n^2} \sum_{i=1}^{n} i X_i$$

is consistent estimator of μ.

Solution :

$$\begin{aligned} E(\bar{X}) &= \frac{2}{n^2} \sum_{i=1}^{n} i E(X_i) \\ &= \frac{2}{n^2} \cdot \frac{n(n+1)}{2} \mu \\ &= \frac{n+1}{n} \mu \end{aligned}$$

$$\begin{aligned} V(\bar{X}) &= \frac{4}{n^4} \sum_{i=1}^{n} i^2 V(x_i) \\ &= \frac{4}{n^4} \cdot \frac{n(n+1)(2n+1)}{6} \cdot \sigma^2 \\ &= \frac{2(n+1)(2n+1)}{3n^3} \sigma^2 \end{aligned}$$

Since

$$\lim_{n \to \infty} E(\bar{X}) = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)\mu = \mu$$

and

$$\lim_{n \to \infty} V(\bar{X}) = \frac{2}{3}\sigma^2 \lim_{n \to \infty} \frac{(n+1)(2n+1)}{n^3} = 0$$

You have by last theorem that X is consistent estimator of μ.

Example 6 :

(Example 4 continued) Show that n min $(X_1, X_2, \ldots, X_n)$ is not consistent for expected time to failure $E(T) = \frac{1}{\alpha}$

Solution :

From solution of example 4 we find that for any $\varepsilon > 0$

$$P\left(\left| nM - \frac{1}{\alpha} \right| < \varepsilon\right) = P\left(\frac{1}{n}\left(\frac{1}{\alpha} - \varepsilon\right) < M < \frac{1}{n}\left(\frac{1}{\alpha} + \varepsilon\right)\right)$$

$$= \int_{\frac{1}{n}\left(\frac{1}{n} - \varepsilon\right)}^{\frac{1}{n}\left(\frac{1}{n} + \varepsilon\right)} n\alpha e^{-n\alpha m}\, dm$$

$$= e^{-\left(\frac{1}{\alpha} - \varepsilon\right)} - e^{-\left(\frac{1}{\alpha} + \varepsilon\right)}$$

which is independent of n.

Thus

$$\lim_{n \to \infty} P\left(\left| nM - \frac{1}{\alpha} \right| < \varepsilon\right) = e^{-\left(\frac{1}{\alpha} - \varepsilon\right)} - e^{-\left(\frac{1}{\alpha} + \varepsilon\right)} \neq 1$$

so nM is not consistent estimator of $E(T) = \frac{1}{\alpha}$.

Check Your Progress B

E1

Suppose population distribution is Poisson with unknown parameter λ. Show that $a\bar{X} + (1 - a)s^2$ is unbiased estimator *for* λ for any $0 \le a \le 1$, a being a constant.

E2

We know s^2 is unbiased for population variance σ^2. Show that s is not unbiased for σ. Hint : Start with $(E(s))^2 = \sigma^2$ and obtain a contradiction.

E3

Is $\overline{X}^2$ unbiased estimator for μ^2, μ being population mean ? If not find its bias.

E4

Consider the class of estimators

$$\overline{X}_a = \sum_{i=1}^{n} a_i X_i.$$

Where a_i are any constants subject to the condition

$$\sum_{i=1}^{n} a_i = 1.$$

Show that sample mean $\overline{X}$ is minimum variance unbiased estimator for estimating population mean μ in this class of estimators.

E5

Show that

$$\bar{X} + \frac{1}{n}$$

is a consistent estimator for population mean μ.

E6

Let $X_1, X_2, ..., X_n$ be a random sample from a uniform distribution on the interval $(0, \theta)$ where θ is unknown parameter. Show that

$$\frac{n+1}{n} \max\ (X_1, ..., X_n)$$

is unbiased estimator of θ. Is it a consistent estimator for parameter θ ?

7.4 METHODS OF POINT ESTIMATION

You can now judge how good is an estimator by applying to it various criteria such as unbiasness, consistency and minimum variance provided an estimator is given to you. Most of the estimators proposed in the previous section were adhoc based on intuitive reasoning. But this may not always work and we should look for some general methods for constructing point estimators of population parameters of interest, clearly we would prefer the methods which are simple to apply and are such that estimators produced by them are reasonable when judged with respect to various criteria of previous section.

You will now learn two most commonly used methods of point estimation viz.

* Method of moments
* Method of maximum likelihood

These methods usually yield reasonable estimators.

7.4.1 Method of Moments

Let the population distribution be $f(x;\theta)$ and assume population parameter $\theta = (\theta_1, \ldots, \theta_k)$ is a vector with k components. You may like to find estimators for all θ_i from a random sample $X_1, X_2, \ldots, X_n$ of size n from this population. The method of moments proceeds as follows.

Step 1 :

Assuming first k moments about origin exist for the population, find them from population distribution $f(x;\theta)$ Let these be

$$\mu'_r(\theta) = E(X^r),\ r = 1, 2, \ldots, k,$$

Evidently they are functions of unknown population parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$.

Step 2 :

Compute first k sample moments about zero

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r, \qquad r = 1, 2, \ldots, k$$

from the random sample.

Step 3 :

Equate population moments to the corresponding sample moments to yield a system of k equations as follows :

$$\mu'_r(\theta) = m'_r, r = 1, 2, \ldots, k$$

Solve this system of equations for k unknowns $\theta_1, \theta_2, \ldots, \theta_k$. The solutions $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ are taken as the estimators of $\theta_1, \theta_2, \ldots, \theta_n$ respectively. Clearly $\hat{\theta}_i$ are functions of sample moments $m'_1, m'_2, \ldots, m'_k$ and hence are statistics.

The estimators obtained by the above method are called moment estimators. Moment estimators may be biased but are consistent.

Note :

The method of moments can not be applied if the population moments do not exist. Since

$$E(m'_r) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^r) = \frac{1}{n}.n\,\mu'_r = \mu'_r$$

therefore the sample moment m'_r is unbiased estimator for the population moment μ'_r. This suggests a basis for Step 3 in the method of moments.

The following examples illustrate the working of method.

Example 7 :

A company assembles 20 tractors per week and ships only those which pass stringent quality tests and the remaining are sent back to assembly line for readjustments. The following data gives number of tractors that failed to quality tests during past 10 weeks.

1 , 0 , 2 , 2 , 1 , 3 , 2 , 0 , 3 , 1

Find the moment estimate of the proportion of tractors that fail to pass quality tests. Is the moment estimator unbiased and consistent ?

Solution :

Let p be the proportion of tractors not passing through the quality tests and X be the number of tractors not passing the quality test during a day. Clearly X has binomial distribution with parameters $n = 20$ and unknown p. Firstly, we must find moment estimator of population parameter p.

Step 1 :

For binomial distribution $\mu'_1 = 20p$.

Step 2 :

Sample moment

$$m'_1 = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} \overline{X}_i \; .$$

Step 3 :

Equating sample and population first moment we have

$$m'_1 = np$$

which given moment estimator of p as

$$\hat{p} = \frac{\overline{X}}{20}.$$

The point estimate of p computed from given data is

$$\hat{p} = \frac{1 + 0 + 2 + \dots + 3 + 1}{10.20}$$

$$= \frac{15}{200} = .075$$

Since $E(\overline{X}) = np$, we have

$$E\left(\frac{\overline{X}}{20}\right) = p$$

and hence $\hat{p}$ is unbiased estimator of p. Also since population variance $\sigma^2 = 20p(1-p)$, we have

$$V(\hat{p}) = V\left(\frac{\overline{X}}{20}\right) = \frac{1}{20^2} \cdot \frac{\sigma^2}{n} = \frac{p(1-p)}{20n}$$

which $\to 0$ as $n \to \infty$. We conclude that $\hat{p}$ is consistent estimator of p.

Example 8 :

Let population distribution be Gamma with pdf

$$f(x; \alpha, \beta) = \frac{1}{\Gamma_\alpha \beta^\alpha} e^{-x/\beta} . x^{\alpha - 1}, x > 0$$

Find moment estimators for parameter α and β based on a random sample of size n from the above population.

Solution :

Step 1 :

Verify that for Gamma distribution

$$\mu'_1 = E(X) = \alpha\beta, \ \mu'_2 = E(\overline{X}^2) = \alpha(\alpha+1)\beta^2$$

Step 2 :

Since there are two parameters to be estimated, compute sample moments

$$m'_1 = \overline{X}$$

and

$$m'_2 = \frac{1}{n}\sum_{i=1}^{n} \overline{X}_i^{\,2}$$

from random sample $\overline{X}_1, \overline{X}_2, ..., \overline{X}_n$.

Step 3 :

Two equations to be solved for α and β are

$$\alpha\beta = m'_1$$

$$\alpha(\alpha+1)\beta^2 = m'_2$$

The solution give the moment estimators of α, β as

$$\hat{\alpha} = m'_1/\hat{\beta}, \ \hat{\beta} = \left(m'_2 - m'^2_1\right)/m'_1.$$

7.4.2 Method of Maximum Likelihood

It is the most widely used method for finding point estimators, since the estimator produced by this method has many nice statistical properties, particularly if sample size is large. Let us first see how does the method work.

We continue with our convention of denoting $f(x;\theta)$ as the population *pdf* or *pmf* as the case may be and θ may be single or vector parameter. The method of maximum likelihood for estimating parameter θ from random sample $X_1, X_2, ..., X_n$ is based on an intuitively appealing principle. Suppose we have observed a random sample, that is $X_1, = x_1, ..., X_n = x_n$ so that $(x_1, x_2, ..., x_n)$ is actually observed data or sample. Then the method of maximum likelihood picks out of all possible values of θ, the one most likely to have produced the observed sample $x_1, x_2, ..., x_n$. Let us explain the principle through a simple illustration, which will also form our basis for writing down steps involved in the method.

In studying the traffic flow at an intersection during peak hours, the data on the random variable X representing the number of vehicles passing through the intersection during an interval of 5 minutes was collected independently on 4 occasions. The data is thus an observed sample of size 4 on X and is given as $x_1 = 100, x_2 = 40, x_3 = 70, x_4 = 54$.

In such studies X is known to follow Poisson distribution with parameter λ. The probability of observing the given sample of size 4 is

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$$

$$= \prod_{i=1}^{4} P(X = x_i)$$

$$= \prod_{i=1}^{4} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-4\lambda} \lambda^{x_1 + x_2 + x_3 + x_4}}{\prod_{i=1}^{4} x_i!}$$

since by definition if random sample X_1, X_2, X_3, X_4 are independent random variable each having Poisson distribution with parameter λ. The last probability is clearly a function of unknown population parameter λ, which we shall denoted by $L(\lambda / x_1, x_2, x_3, x_4)$ or simply by $L(\lambda)$. This function will be called the likelihood function of parameter λ given observed sample (x_1, x_2, x_3, x_4).

The value of λ that makes the likelihood function maximum is infact the value which gives the highest probability of observing the sample that we did observe. This value of λ say $\hat{\lambda}$ is called the likelyhood estimator of λ, based on observed sample (x_1, x_2, x_3, x_4). One can use the method calculus to maximise likelihood function $L(\lambda)$. Since natural logarithm is increasing function, λ that maximises $L(\lambda)$ also maximises $\ln L(\lambda)$. It will be computationally simpler to maximise

$$\ln L(\lambda) = -4\lambda + \sum_{i=1}^{4} x_i (n\lambda - \ln \prod_{i=1}^{4} x_i).$$

Differentiating $\ln L(\lambda)$ and equating the derivative to zero we get

$$\frac{d}{d\lambda} \ln L(\lambda) = -4 + \frac{\sum_{i=1}^{4} x_i}{\lambda} = 0$$

The solution $\hat{\lambda}$ of λ in the last equation is

$$\hat{\lambda} = \frac{\sum_{i=1}^{4} x_i}{4} = \frac{100 + 40 + 70 + 54}{4} = 66$$

Since

$$\frac{d^2}{d\lambda^2} \ln L(\lambda) = -\frac{\sum Xi}{\lambda^2} < 0 \text{ for } \lambda = \hat{\lambda}$$

We see that $\ln L(\lambda)$ has a maximum at $\hat{\lambda}$. Thus the likelihood estimate of λ based on given observd sample is

$$\hat{\lambda} = \bar{x} = 66$$

This is just a value taken by statistic

$$\bar{X} = \frac{1}{4} \sum_{i=1}^{4} X_i$$

$\bar{X}$ is called likelihood estimator of parameter λ.

From the above discussion you are now in a position to write the main steps involved in finding maximum likelihood estimator for a single parameter θ. These are

Step 1 :

Observe a sample ($x_1, x_2, \ldots\ldots, x_n$) from probability distribution $f(x; \theta)$ and compute the likelihood function of θ given the sample ($x_1, x_2, \ldots\ldots, x_n$) which is

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

$$= \prod_{i=1}^{n} f(x_i; \theta)$$

Step 2 :

Find the admissible value of θ, say $\hat{\theta}$ for which $L(\theta)$ is maximum. That is

$$\max L(\theta) = L(\hat{\theta}).$$

If $L(\theta)$ is a differentiable function, you can find this value of θ that maximises ln $L(\theta)$ by solving the equation

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

for θ. This equation is called *likelihood equation.* The likelihood estimate $\hat{\theta}$ of θ is thus solution of likelihood equation.

Step 3 :

Let $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots\ldots, x_n)$ be the likelihood estimate as obtained in Step 2. Replace ($x_1, x_2, \ldots\ldots, x_n$) in $\hat{\theta}$ by random sample ($X_1, X_2, \ldots\ldots, X_n$) to obtain the maximum likelihood estimator (MLE) $\hat{\theta}(X_1, X_2, \ldots, X_n)$ of parameter θ.

It may happen in some problem that population distribution depends on more than one parameter. In that case θ is a vector, say $\theta = (\theta_1, \theta_2, \ldots\ldots, \theta_k)$ then in Step 2 the likelihood equation is replaced by system of simultaneous equations

$$\frac{\partial}{\partial \theta_i} \ln L(\theta) = 0, \; i = 1, 2, \ldots, k$$

in k unknowns $\theta_1, \theta_2, \ldots, \theta_k$.

The Step 2 can be performed only if derivatives exist, otherwise one has to look for alternative method for maximization of likelihood function.

The likelihood equation in most practical problems has a unique solution and the solution maximises likelihood function. So the second derivative test for maximum can be omitted.

Properties of MLE

Some of the general properties of MLE are stated below without proofs.

1. The MLE may not be unbiased.
2. If $\hat{\theta}$ is MLE of θ, then MLE of $h(\theta)$ a function of θ is $h(\hat{\theta})$. This property is referred to as invariance property of MLE.

3. MLE is consistent under fairly general regularity condition on population distribution.

4. Let $f(x;\theta)$ be *pdf* (or *pmf*) of the population and θ be a real parameter. Let $\hat{\theta}$ be MLE of θ based on a random sample of size n. If n is sufficiently large, then the probability distribution of $\hat{\theta}$ is approximately

$$N\left(\theta, \frac{1}{I(\theta)}\right)$$

where

$$I(\theta) = nE\left[\frac{d}{d\theta}\ln f(X;\theta)\right]^2$$

Under some general regularity conditions on f.

Most of the common probability distributions satisfy the regularity conditions needed for properties 3 and 4 and you should not bother at this introductory level of study of the subject.

After studying the following examples you should be able to apply the method of maximum likehood estimator to construct MLE for paranfeter of you interest.

Example 9 :

In the manufacture of a certain type of item on a machine, it is observed that not all manufactured items conform to the specifications due to uncontrolable factors. The manufacture would like to estimate proportion of defective items produced by the machine. A sample of n items selected randomly from production line was tested and r items were found to be defective. Find MLE of p using this data. Show that MLE of p is unbiased and consistent for p. Write approximate probability distribution of MLE if $n = 120$.

Solution :

With any item manufactured by the machine we associate a random variable X defined by

$$X = \begin{cases} 1, & \text{if item is defective} \\ 0, & \text{if item is not defective} \end{cases}$$

The X has Bernoulli's distribution

$$f(x;p) = P(X = x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

which is taken to be the population distribution. We have to find MLE of population parameter p.

Step 1 :

Let X_i be the value of X for i th, $i = 1, 2, \ldots, n$ item chosen in the sample. Then clearly $X_1 + X_2 + \ldots + X_n = r$, the number of defective in the sample.

The likelihood function

$$L(p/x_1, x_2, \ldots, x_n) = L(p) \text{ is}$$

$$L(p) = \prod_{i=1}^{n} f(x_i;p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum x_i} = p^r(1-p)^{n-r}$$

Step 2 :

The likelihood equation is

$$\frac{d}{dp}\ln L(p) = \frac{r}{p} - \frac{n-r}{1-p} = 0$$

The solution of this equation is say $\hat{p}$

Then

$$\hat{p} = \frac{r}{n} = \bar{x}$$

Step 3 :

MLE of p is sample mean $\bar{X}$, or the sample proportion defective.

Since r has binomial distribution with parameters n and p, we have

$$E(r) = np, \quad V(r) = np(1-p)$$

or

$$E(\hat{p}) = p$$

and

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

Thus MLE $\hat{p}$ of p is unbiased and since

$$V(\hat{p}) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

it is also consistent for p.

Since sample size n is large, using property of MLE we have approximate distribution of $\hat{p}$ as normal

$$N\left(p, \frac{1}{I(p)}\right),$$

where

$$I(p) = nE\left[\frac{d}{dp}\ln f(X;p)\right]^2$$

$$= nE\left[\frac{d}{dp}(X\ln p + (1-X)\ln(1-p))\right]^2$$

$$= nE\left(\frac{X}{p} - \frac{1-X}{1-p}\right)^2$$

$$= nE\left(\frac{X-p}{p(1-p)}\right)^2 = \frac{n}{p^2(1-p)^2}V(X) = \frac{np(1-p)}{p^2(1-p)^2}$$

$$= \frac{n}{p(1-p)}$$

Hence $\hat{p}$ has approximately normal distribution

$$N\left(p, \frac{p(1-p)}{n}\right)$$

Example 10 :

A doze of drug causes loss of sleep in human beings. Loss of sleep, say X hours is a random time period anywhere between 0 and θ hours, θ being unknown. A medical researcher will be interested in estimating θ. Each of a sample of n randomly chosen persons was given a doze of drug and the loss of sleep was observed for each of these n persons as $X_1, X_2, \ldots, X_n$. Find MLE of θ based on this sample. Is MLE of θ unbiased and consistent?

Also find MLE of mean loss of sleep.

Solution :

The loss of sleep X hours is a random variable having a uniform distribution on the interval $(0, \theta)$. Thus the population *pdf* is

$$f(X;\theta) = \frac{1}{\theta},\quad 0 \le x \le \theta, \quad = 0 \text{ otherwise}$$

Step 1 :

Likelihood function is given by

$$L(\theta) = L(\theta/x_1, \ldots, x_n) = \left(\frac{1}{\theta}\right)^n \quad 0 \le x_i \le \theta, i = 1, 2, \ldots, n$$

$$= 0 \quad \text{otherwise.}$$

$$\text{equivalently } L(\theta) = \frac{1}{\theta^n},\quad \max_{1 \le i \le n} x_i \le \theta$$

$$= 0, \text{ otherwise}$$

Step 2 :

Since $L(\theta)$ is not differentiable at $\theta = \max\limits_{1 \le i \le n} x_i$, the method of calculus fails.
However from the graph (see Figure 7.3) of $L(\theta)$, it is clear that maximum of $L(\theta)$ occurs at

$$\theta = \hat{\theta} = \max_{1 \le i \le n} x_i$$

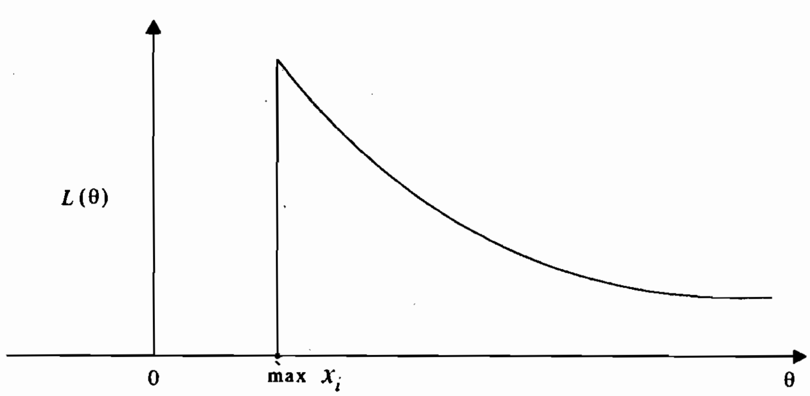

Figure 7.3

Step 3 :

MLE of θ is maximum observation that is ,

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \max_{1 \le i \le n} X_i$$

Let us now determine mean and variance of $\hat{\theta}$. The distribution function of

$$M = \max_{1 \le i \le n} X_i$$

is given by

$$P(M \le x) = P\left(\max_{1 \le i \le n} X_i \le x\right)$$

$$= P(X_1 \le x, X_2 \le x, \ldots, X_n \le x)$$

$$= \prod_{i=1}^{n} P(X_i \le x)$$

$$= \left(\frac{x}{\theta}\right)^n, \; 0 \le x \le \theta$$

and $=$ 1, for $x > \theta$. This is because each X_i has independent uniform distribution on the interval $(0, \theta)$.

Thus *pdf* of M is

$$f(x) = \frac{n}{\theta^n} x^{n-1}, \quad 0 \le x \le \theta$$

$$= 0, \text{ otherwise}$$

and

$$E(M^i) = \int_0^\theta x^i \frac{n}{\theta^n} x^{n-1}\, dx = \frac{n\,\theta^i}{n+i} \qquad i = 1, 2, \ldots$$

From which we get

$$E(M) = \frac{n}{n+1}\,\theta$$

and

$$V(M) = \frac{n}{n+2}\,\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2$$

$$= \frac{n}{(n+1)^2(n+2)}\,\theta^2$$

Since $E(M) \to \theta$, and $V(M) \to 0$ as $n \to \infty$, we conclude that M or $\hat{\theta}$ is consistent for θ, but it is biased with bias

$$B(\theta) = -\frac{\theta}{n+1}$$

Mean loss of sleep is

$$E(X) = \frac{\theta}{2}$$

By invariance property of MLE we conclude that MLE of $E(X)$ is

$$\frac{1}{2}\,\hat{\theta} \text{ or } \frac{M}{2}.$$

Example 11 :

Find MLE of parameters μ and σ^2 of normal distribution $N(\mu, \sigma)^2$, based on a random sample $(X_1, \ldots, X_n)$ of size n from this distribution. Also find MLE of $\frac{\sigma}{\mu}$.

Solution :

Step 1 :

Likelihood function is given by

$$L(\mu, \sigma^2) = L(\mu, \sigma^2 / x_i, \ldots, x_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

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

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2}\ln(2\pi)$$

Step 2 :

The system of likelihood equations are

$$\frac{\partial}{\partial \mu}\ln L(\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} -2(x_i-\mu) = 0$$

$$\frac{\partial}{\partial(\sigma^2)}\ln L(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

Solving these equations for μ, σ^2 we get maximum likelihood estimates as

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

Step 3 :

The MLE for μ and σ^2 are $\hat{\mu} = \bar{X}$ the sample mean and

$$\hat{\sigma}^2 = \frac{n-1}{n} s^2,$$

where s^2 is the sample variance.

Using invariance property of MLE we get the MLE of $\frac{\sigma}{\mu}$

as

$$\frac{\hat{\sigma}}{\hat{\mu}} = \sqrt{\frac{n-1}{n}}\,\frac{s}{\bar{X}}$$

In the examples considered above the likelihood equations could be easily solved to obtain explicit likelihood estimates of the population parameters. But this may not be possible always. Just set up the likelihood equations for the two parameters of gamma distribution and try to solve it. Do you get explicit solutions ? You will fail in this.

Check Your Progress C

E1

Let the population *pdf* be

$$f(x;\theta = (\theta+1)x^{\theta}, 0 < x < 1$$

Based on a random sample of size n, find (a) moment estimator (b) MLE of parameter θ. If an observed sample is obtained as 0.35 , 0.72 , 0.50 , 0.31 , 0.80, what are moment and maximum likelihood estimates of θ.

E2

A certain type of electric tube is required to be used for 1500 hours. Assume life length of the tube has exponential distribution with unknown mean θ. Find MLE of the probability that a tube of this type gives a failure free operation for at least 1500 hrs. What is the estimate of this probability if sample of 6 tubes when tested gave the following life length observations.

1100 , 1085 , 1585 , 1602 , 1540 , 1750

E3

The life length T of a component has *pdf* $f(t) = \alpha e^{-\alpha(t-\beta)}, t > \beta > 0$. Based on a random sample of size n on T find MLE of

(a) α if β is known.

(b) β if α is known.

E4

A random sample of size 50 is taken from normal distribution $N(0, \sigma^2)$. Find MLE of standard deviation σ. Assuming 50 to be a large sample size, find approximately the probability

$$P\left(0.9 < \frac{\hat{\sigma}}{\sigma} < 1.1\right),$$

where $\hat{\sigma}$ is MLE of σ.

E5

A random sample of size n is taken from geometric distribution with parameter p. Find moment estimator and MLE for p. Are they same ?

7.5 SUMMARY

In this unit you have learnt

a) the standard statistical terminology

b) the common characteristics of problems needing statistial study

c) the basic ideas in the theory of point estimation

d) that the performance of a point estimator can be judged with respect to criterion of unbiasedness, consistency and minimum variance

e) that the sample mean $\overline{X}$ and sample variance s^2 are good estimators for population mean μ and population variance σ^2 respectively

f) the method of moments for constructing point estimator

g) the principle of maximum likehood and the method of maximum likelihood for constructing point estimators

h) the properties of maximum likelihood estimators

i) the use and derivation of the asymptotic distribution of maximum likelihood estimator in case of large sample sizes.

7.6 SOLUTIONS/ANSWERS

Check Your Progress A

E1

(i) Collection of all workers in the factory

(ii) $X = \begin{cases} 0, & \text{worker does not favour new pension scheme} \\ 1, & \text{worker favours now pension scheme} \end{cases}$

(iii) p = probability or proportion of workers favouring the new scheme.

(iv) $P(X = x) = p^x (1-p)^{1-x}, x = 0, 1.$

(v) S_p = proportion of workers favouring new scheme in a sample of size n.

$= (X_1 + \ldots\ldots + X_n)/n$

where $(X_1, \ldots, X_n)$ is random sample of size n.

E2

(i) Collection of all farmers in the region

(ii) Normal $N(\mu, \sigma^2)$

(iii) μ = mean yield / acre

(iv) Sample mean $\overline{X}$

(v) 3677.9

(vi) 2122881

Check Your Progress B

E1

For Poisson distribution mean $\mu = \lambda$ and variance $\sigma^2 = \lambda$. Thus using theorem 1 and 2 we have $E(\overline{X}) = \lambda$, $E(s^2) = \lambda$ and therefore $E(a\overline{X} + (1-a)s^2) = a\lambda + (1-a)\lambda = \lambda.$

E2

If $E(s) = \sigma$ then $(E(s))^2 = \sigma^2$. Also $E(s^2) = \sigma^2$. Hence $E(s)^2 - (E(s)^2) = 0$ or Var $s = 0$, which is not true. Hence a contradiction.

E3

Since $V(\overline{X}) = E(\overline{X}^2) - (E(\overline{X}))^2 = \frac{\sigma^2}{n}$, wehave $E(\overline{X}^2) = \frac{\sigma^2}{n} + \mu^2$.

Thus $\overline{X}^2$ is not unbiased for μ^2. Bias is $\frac{\sigma^2}{n}$.

E4

$$E(\overline{X}_a) = \sum_{i=1}^{n} a_i E(X_i) = \mu \sum_{i=1}^{n} a_i = \mu \Rightarrow \overline{X}_a$$

is unbiased for any a_i's.

$$V(\overline{X}_a) = \sum_{i=1}^{n} a_i^2 V(x_i) = \sigma^2 \sum_{i=1}^{n} a_i^2$$

Also

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \left(a_i - \frac{1}{n}\right)^2 + \frac{1}{n},$$

which has a minimum value when

$$a_i = \frac{1}{n}$$

for all i. Hence $V(\overline{X}_a)$ is minimum for $\overline{X}$

E5

$$E\left(\overline{X} + \frac{1}{n}\right) = \mu + \frac{1}{n} \rightarrow \mu \text{ as } n \rightarrow \infty$$

$$V\left(\overline{X} + \frac{1}{n}\right) = V(\overline{X}) = \frac{\sigma^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Application of theorem 3 gives the desired result.

E6

The *pdf* of $M = \max(X_1, \ldots, X_n)$ is

$$f(x) = \frac{n}{\theta^n} x^{n-1}, 0 \leq x \leq \theta,$$

and = 0 , otherwise

(See Example 10 in Section 7.4)

$$E\left(\frac{n+1}{n} M\right) = \int_o^\theta \frac{n+1}{n} x \,.\, \frac{n}{\theta^n} x^{n-1} dx = \theta$$

$$V\left(\frac{n+1}{n} M\right) = \int_o^\theta \left(\frac{n+1}{n} x\right)^2 \frac{n}{\theta^n} x^{n-1} dx - \theta^2$$

$$= \frac{\theta^2}{n(n+2)} \rightarrow 0, \text{ as } n \rightarrow \infty$$

Thus $\frac{n+1}{n} M$ is unbiased for θ and by Theorem 3 it is also consistent estimator for θ.

Check Your Progress C

E1

(a) $$E(X) = (\theta + 1) \int_0^1 x \,.\, x^\theta dx = \frac{\theta + 1}{\theta + 2}$$

Moment estimator $\hat{\theta}_m$ is solution of

$$\frac{\hat{\theta}_m + 1}{\hat{\theta}_m + 2} = \overline{X},$$

which gives

$$\hat{\theta}_m = \frac{2\overline{X} - 1}{1 - \overline{X}}.$$

(b)

$$\ln L(\theta) = n \ln(1 + \theta) + \theta \sum_{i=1}^{n} \ln x_i$$

$$\frac{d}{d\theta} \ln L(\theta) = \frac{1}{1+\theta} + \sum_{i=1}^{n} \ln x_i = 0$$

Likelihood estimator

$$\hat{\theta}_l = \frac{n}{\sum_{i=1}^{n} \ln \frac{1}{x_i}} - 1$$

(c) $\hat{\theta}_m = 0.1552$, $\hat{\theta}_l = 0.4426$.

E2

Probability that life length ≥ 1500 hours is $e^{-\frac{1500}{\theta}}$

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

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

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i = 0$$

MLE of θ is

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

Therefore MLE of

$e^{-\frac{1500}{\theta}}$ is $e^{-\frac{1500}{\overline{X}}}$

The estimate of $e^{-\frac{1500}{\theta}}$ based on data is 0.3538.

E3

$$L(\alpha, \beta) = \alpha^n e^{-\alpha \sum_{i=1}^{n} (t_i - \beta)}, \; t_i \geq \beta, \beta \text{ known}$$

(a) $$\frac{\partial}{\partial \alpha} \ln L = \frac{n}{\alpha} - \sum_{i=1}^{n} (t_i - \beta) = 0$$

$$\Rightarrow \text{MLE } \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} (T_i - \beta)}.$$

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

$\min_{1 \le i \le n} t_i \ge \beta$, α known

$L (\alpha, \beta)$ is maximum at maximum value of β which is

$\min_{1 \le i \le n} t_i$. Thus MLE of β is $\hat{\beta} = \min (T_1, \ldots, T_n)$.

E4

$$L (\sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2 \pi} \sigma} e^{- x_i^2 / 2 \sigma^2},$$

where $n = 50$.

$$\frac{d}{d\sigma} \ln L (\sigma) = - \frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \text{MLE of } \sigma \text{ is } \hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} X_i^2}$$

Since

$$f(x ; \sigma) = \frac{1}{\sqrt{2 \pi} \sigma} e^{- \frac{1}{2} \frac{x^2}{\sigma^2}}$$

is the population *pdf*, we have

$$\frac{d}{d\sigma} \ln f(x ; \sigma) = - \frac{1}{\sigma} + \frac{1}{\sigma^3} X^2 = \frac{X^2 - \sigma^2}{\sigma^3}$$

and

$$I (\sigma) = nE \left(\frac{X^2 - \sigma^2}{\sigma^3} \right)^2 = \frac{n}{\sigma^6} \left[E (X^4) - 2 \sigma^2 E (X^2) + \sigma^4 \right]$$

For $N (0, \sigma^2)$ distribution $E (X^4) = 3 \sigma^4, E (X^2) = \sigma^2$

and, therefore, $I (\sigma) = 2n/\sigma^2$

Thus $\hat{\sigma}$ has approximately a normal distribution

$$N \left(\sigma, \frac{\sigma^2}{2n} \right)$$

for large n. Therefore with $n = 50$ we have

$$P\left(0.9 < \frac{\hat{\sigma}}{\sigma} < 1.1\right) = P\left(\left|\frac{\hat{\sigma} - a}{\sigma/\sqrt{100}}\right| < 1\right)$$

$$= 2\Phi(1) - 1 = 0.6826$$

(from normal distribution tables).

E5

For geometric distribution

$$E(X) = \frac{1}{p}$$

Moment estimator $\hat{p}_m$ is given by

$$\frac{1}{\hat{p}_m} = \overline{X}.$$

Hence

$$\hat{p}_m = \frac{1}{\overline{X}}$$

$$L(p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{\sum_{i=1}^{n} x_i - n} p^n$$

$$\frac{d}{dp} \ln L(p) = \frac{n}{p} - \frac{\left(\sum_{i=1}^{n} x_i - n\right)}{1-p} = 0$$

Solving we get MLE $\hat{p}_l$ as

$$\hat{p}_l = \frac{1}{\overline{X}}$$

Thus moment estimator and MLE are the same.