UNIT 5 INTRODUCTION TO ESTIMATION

Structure

5.1 INTRODUCTION

In many real-life problems, the population parameter(s) is (are) unknown and someone is interested to obtain the value(s) of parameter(s). But, if the whole population is too large to study or the units of the population are destructive in nature or there is a limited resources and manpower available then it is not practically convenient to examine each and every unit of the population to find the value(s) of parameter(s). In such situations, one can draw sample from the population under study and utilize sample observations to estimate the parameter(s).

Every one of us makes estimate(s) in our day to day life. For example, a house wife estimates the monthly expenditure on the basis of particular needs, a sweet shopkeeper estimates the sale of sweets on a day, etc. So the technique of finding an estimator to produce an estimate of the unknown parameter on the basis of a sample is called estimation.

There are two methods of estimation:

1. Point Estimation

2. Interval Estimation

In point estimation, we determine a appropriate single statistic whose value is used to estimate the unknown parameter whereas in interval estimation, we determine an interval that contains true value of the unknown parameter with certain confidence. The point estimation and interval estimation are briefly described in Unit 6 and Unit 7 respectively of this block.

Estimation admits two problems; the first is to select some criteria or properties such that if an estimator possesses these properties it is said to be the best estimator among possible estimators and the second is to derive some methods or techniques through which we obtain an estimator which possesses such properties. This unit is devoted to explain the criteria of good estimator. This unit is divided into nine sections. Section 5.1 is introductory in nature. The basic terms used in estimation are defined in Section 5.2. Section 5.3 is devoted to criteria of good estimator which are explained one by one in subsequent sections. Section 5.4 is explored the concept of unbiasedness with examples. Unbiasedness is based on fixed sample size whereas the concept based on varying sample size, that is, consistency is described in Section 5.5. There exists more than one consistent estimator of a parameter, therefore in Section 5.6 is explained the next property efficiency. Section 5.7 is devoted to describe sufficiency. Unit ends by providing summary of what we have discussed in this unit in Section 5.8 and solution of exercises in Section 5.9.

Objectives

After studying this unit, you should be able to:

- define the parameter space and joint probability density (mass) function;
- describe the characteristics of an estimator;
- explain the unbiasedness of an estimator;
- explain the consistency of an estimator;
- explain the efficiency of an estimator;
- explain the most efficient estimator;
- explain the sufficiency of an estimator; and
- describe the minimum variance unbiased estimator.

5.2 BASIC TERMINOLOGY

Before discussing the properties of a good estimator, we discuss basic definitions of some important terms. These terms are very useful in understanding the fundamentals of theory of estimation discussed in this block.

Discrete and Continuous Distributions

In Units 12 and 13 of MST-003, we have discussed standard discrete and continuous distributions as binomial, Poisson, normal, exponential, etc. We know that the populations can be described with the help of distributions, therefore, standard discrete and continuous distributions are used in statistical inference. Here, we discuss some standard discrete and continuous distributions in brief as in tabular form:

S. No.	Distribution	Parameter(s)	Mean	Variance
1	Bernoulli (discrete) $P[X=x]=p^x(1-p)^{1-x};\ \ x=0,1$	p	p	pq
2	Binomial (discrete) $P[X=x]={}^nC_x p^x q^{n-x};\ \ x=0,1,\ldots,n$	n & P	np	npq
3	Poisson (discrete) $P[X=x]=\dfrac{e^{-\lambda}\lambda^x}{x!};\ \ x=0,1,\ldots\ \&\ \lambda>0$	λ	λ	λ
4	Uniform (discrete) $P[X=x]=\dfrac{1}{n};\ \ x=0,1,\ldots,n$	n	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$
5	Hypergeometric (discrete) $P[X=x]=\dfrac{{}^MC_x\ {}^{N-M}C_{n-x}}{{}^NC_n};\ x=0,1,\ldots,\min\{M,n\}$	N, M & n	$\dfrac{nM}{N}$	$\dfrac{NM(N-M)(N-n)}{N^2(N-1)}$

6	Geometric (discrete) $P[X=x]=pq^x;\ \ x=0,1,2,...$	p	$\frac{p}{q}$	$\frac{p}{q^2}$
7	Negative Binomial (discrete) $P[X=x]={}^{x+r-1}C_{r-1}p^r q^x;\ \ x=0,1,2,\ldots$	r & p	$\frac{rp}{q}$	$\frac{rp}{q^2}$
8	Normal(continuous) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};\ -\infty<x<\infty$ $\&\ \sigma>0,\ -\infty<\mu<\infty$	μ & σ^2	μ	σ^2
9	Standard Normal(continuous) $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2};\ \ \ -\infty<x<\infty$	--	0	1
10	Uniform (continuous) $f(x)=\frac{1}{b-a};\ a<x<b, b>a$	a & b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
11	Exponential (continuous) $f(x)=\theta e^{-\theta x};\ \ \ x\geq 0\ \&\ \theta>0$	θ	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$
12	Gamma (continuous) $f(x)=\frac{a^b}{\lceil b}e^{-ax}x^{b-1};\ x>0\ \&\ a>0$	a & b	$\frac{b}{a}$	$\frac{b}{a^2}$
13	Beta First Kind (continuous) $f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1};\ 0<x<1$ $\&\ a>0, b>0$	a & b	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
14	Beta Second Kind (continuous) $f(x)=\frac{1}{B(a,b)}\frac{x^{a-1}}{(1+x)^{a+b}};\ x>0$ $\&\ a>0, b>0$	a & b	$\frac{a}{b-1}$	$\frac{a(a+b+1)}{(b-1)^2(b-2)}$

Parameter Space

The set of all possible values that the parameter θ or parameters $\theta_1, \theta_2, \ldots, \theta_k$ can assume is called the parameter space. It is denoted by Θ and is read as **"big theta"**. For example, if parameter θ represents the average life of electric bulbs manufactured by a company then parameter space of θ is $\Theta=\{\theta:\theta\geq 0\}$, that is, the parameter average life θ can take all possible values greater than or equal to 0, Similarly, in normal distribution (μ, σ^2), the parameter space of parameters μ and σ^2 is $\Theta=\{(\mu,\sigma^2):-\infty<\mu<\infty;\ \ 0<\sigma<\infty\}$.

Joint Probability Density (Mass) Function

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population whose probability density (mass) function is $f(x,\theta)$ where, θ is the population parameter then the joint probability density (mass) function of sample values is denoted by $f(x_1, x_1, \ldots, x_n, \theta)$ and defined as

For discrete case,

$$f(x_1, x_1, \ldots, x_n, \theta)=P[X_1=x_1, X_2=x_2, \ldots, X_n=x_n]$$

since $X_1, X_2, ..., X_n$ are independent, therefore,

$$f\left(x_1, x_1, ..., x_n, \theta\right) = P\left[X_1 = x_1\right] P\left[X_2 = x_2\right] ... P\left[X_n = x_n\right]$$

In this case, the function $f\left(x_1, x_1, ..., x_n, \theta\right)$ represents the probability that the particular sample $x_1, x_2, ..., x_n$ has been drawn for a fixed (given) value of parameter θ.

For continuous case,

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

In this case, the function $f\left(x_1, x_1, ..., x_n, \theta\right)$ represents the probability density function of the random sample $X_1, X_2, ..., X_n$.

The process of finding the joint probability density (mass) function is described by taking some examples as:

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

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

then joint probability mass function of $X_1, X_2, ..., X_n$ can be obtained as

$$f\left(x_1, x_1, ..., x_n, \lambda\right) = P\left[X_1 = x_1\right] P\left[X_2 = x_2\right] ... P\left[X_n = x_n\right]$$

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

$$= \frac{e^{\overbrace{-\lambda-\lambda-...-\lambda}^{n-\text{times}}} \lambda^{x_1+x_2+...+x_n}}{x_1!x_2!...x_n!}$$

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

Similarly, if $X_1, X_2, ..., X_n$ is a random sample of size n taken from exponential population whose pdf is given by

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

then the joint probability density function of sample values can be obtained as

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

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

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

$$f\left(x_1, x_1, ..., x_n, \theta\right) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

Let us check your understanding of above by answering the following exercises.

E1) What is the pmf of Poisson distribution with parameter $\lambda = 5$. Also find the mean and variance of this distribution.

E2) If θ represents the average marks of IGNOU's students in a paper of 50 marks. Find the parameter space of θ.

E3) A random sample $X_1, X_2, ..., X_n$ of size n is taken from Poisson distribution whose pdf is given by

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

Obtain joint probability mass function of $X_1, X_2, ..., X_n$.

5.3 CHARACTERISTICS OF ESTIMATORS

It is to be noted that a large number of estimators can be proposed for an unknown parameter. For example, if we want to estimate the average income of the persons living in a city then the sample mean, sample median, sample mode, etc. can be used to estimate the average income. Now, the question arises, "Are some of possible estimators better, in some sense, than the others?" Generally, an estimator can be called good for two different situations:

(i) **When the true value of parameter is being estimated is known−** An estimator might be called good if its value close to the true value of the parameter to be estimated. In other words, the estimator whose sampling distribution concentrates as closely as possible near the true value of the parameter may be regarded as the good estimator.

(ii) **When the true value of the parameter is unknown−** An estimator may be called good if the data give good reason to believe that the estimate will be closed to the true value.

In the whole estimation, we estimate the parameter when the true value of the parameter is unknown. Hence, we must choose estimates not because they are certainly close to the true value, but because there is a good reason to believe that the estimated value will be close to the true value of parameter. In this unit, we shall describe certain properties, which help us in deciding whether an estimator is better than others.

Prof. Ronald. A. Fisher, a Great English Mathematical Statistician. (1890-1962)

Prof. Ronald A. Fisher was the man who pushed ahead the theory of estimation and introduced these concepts and gave some properties of good estimator as follows:

1. Unbiasedness
2. Consistency
3. Efficiency
4. Sufficiency

We shall discuss these properties one by one in the subsequent sections.

Now, give the answer of the following exercise.

E4) Write the four properties of good estimator.

5.4 UNBIASEDNESS

Generally, population parameter(s) is (are) unknown and if the whole population is too large to study to find the value of unknown parameter(s) then one can estimate the population parameter(s) with the help of estimator(s) which is(are) always a function of sample values.

An estimator is said to be unbiased if the expected value of the estimator is equal to the true value of the parameter being estimated.

An estimator is said to be unbiased for the population parameter such as population mean, population variance, population proportion, etc. if and only if the average or mean of the sampling distribution of the estimator is equal to the true value of the parameter.

Mathematically,

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population whose probability density (mass) function is f(x,θ) where, θ is the population parameter then an estimator T = t($X_1, X_2, ..., X_n$) is said to be unbiased estimator of the parameter θ if and only if

$$E(T) = \theta; \quad \text{for all } \theta \in \Theta$$

This property of estimator is called unbiasedness.

Normally, it is preferable that the expected value of the estimator should be exactly equal to the true value of the parameter being estimated. But if the expected value of the estimator does not equal to the true value of parameter, then the estimator is said to be "biased estimator", that is, if

$$E(T) \neq \theta$$

then estimator T is called biased estimator of θ.

The amount of biasness is given by

$$b(\theta) = E(T) - \theta$$

If b(θ) > 0 or E(T) > θ, then the estimator T is said to be positively biased for parameter θ.

If b(θ) < 0 or E(T) < θ, then the estimator T is said to be negatively biased for parameter θ.

If $E(T) \to \theta$ as $n \to \infty$ i.e. if an estimator T is unbiased for a large sample only then estimator T is said to be asymptotically unbiased for θ.

Now, we explain the procedure how to show that a statistic is unbiased or not for a parameter with the help of some examples:

Example 1: Show that sample mean $(\overline{X})$ is an unbiased estimator of the population mean (μ) if it exists.

Solution: Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from any population with mean μ. Then for unbiasedness we have to show that

$$E\left(\overline{X}\right) = \mu$$

Consider,

$$E\left(\overline{X}\right) = E\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] \qquad \left[\text{By defination of sample mean}\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \left[\because E(aX+bY) = aE(X)+bE(Y)\right]$$

Since X_1, X_2,…,X_n are randomly drawn from same population so they also follow the same distribution as the population. Therefore,

$$E(X_1) = E(X_2) = ... = E(X_n) = E(X) = \mu$$

Thus,

$$E(\bar{X}) = \frac{1}{n}\left(\underbrace{\mu + \mu + ... + \mu}_{n-\text{times}}\right)$$

$$= \frac{1}{n}(n\mu) = \mu$$

$$E(\bar{X}) = \mu$$

Hence, sample mean $(\bar{X})$ is an unbiased estimator of the population mean μ.

Also if $x_1, x_2, ..., x_n$ are the observed values of the random sample $X_1, X_2, ..., X_n$ then $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is unbiased estimate of population mean.

If X and Y are two random variables and a & b are two constants then by the addition theorem of expectation, we have $E(aX+bY) = aE(X)+bE(Y)$

Example 2: A random sample of 10 cadets of a centre is selected and measures their weights (in kg) which are given below:

48, 50, 62, 75, 80, 60, 70, 56, 52, 78

Determine an unbiased estimate of the average weight of cadets of the centre.

Solution: We know that sample mean $(\bar{X})$ is an unbiased estimator of the population mean and its particular value is the unbiased estimate of population mean, therefore,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + ... + X_n}{n}$$

$$= \frac{48+50+62+75+80+60+70+56+52+78}{10} = 63.10$$

Hence, an unbiased estimate of the average weight of cadets of the centre is 63.10 kg.

Example 3: A random sample $X_1, X_2, ..., X_n$ of size n taken from a population whose pdf is given by

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

Show that sample mean $(\bar{X})$ is an unbiased estimator of parameter θ.

Solution: For unbiasedness, we have to show that

$$E(\bar{X}) = \theta$$

Here, we are given that

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

Since we do not know the mean of this distribution therefore first of all we find the mean of this distribution. So we consider,

$$E(X) = \int_0^\infty x f(x, \theta) dx$$

$$= \int_0^\infty x \frac{1}{\theta} e^{-x/\theta} dx = \frac{1}{\theta} \int_0^\infty x e^{-x/\theta} dx$$

$$= \frac{1}{\theta} \int_0^\infty x^{2-1} e^{-x/\theta} dx \quad = \frac{1}{\theta} \times \frac{\overline{|2}}{(1/\theta)^2} = \theta \qquad \left[\because \int_0^\infty x^{n-1} e^{-ax} dx = \frac{\overline{|n}}{a^n}\right]$$

Since $X_1, X_2, \ldots, X_n$ are randomly drawn from same population having mean θ, therefore,

$$E(X_1) = E(X_2) = \ldots = E(X_n) = E(X) = \theta$$

Consider,

If X and Y are two random variables and a & b are two constants then by the addition theorem of expectation, we have $E(aX + bY) = aE(X) + bE(Y)$

$$E(\overline{X}) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] \qquad [\text{By defination of sample mean}]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right] \quad \left[\begin{array}{l} \because E(aX + bY) \\ = aE(X) + bE(Y) \end{array}\right]$$

$$= \frac{1}{n}(\underbrace{\theta + \theta + \ldots + \theta}_{n-\text{times}})$$

$$= \frac{1}{n}(n\theta) = \theta$$

Thus, $\overline{X}$ is an unbiased estimator of θ.

Note 1: If $X_1, X_2, \ldots, X_n$ is a random sample taken from a population with mean μ and variance σ^2, then

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{ is a biased estimator of } \sigma^2.$$

whereas, $S^2 = \frac{n}{n-1} S'^2$ i.e. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an unbiased estimator of σ^2.

The proof of the above result is beyond the scope of this course. But for your convenience, we will show this result with the help of the following example.

Example 4: Consider a population comprising three televisions of certain company. If lives of televisions are 8, 6 and 10 years then construct the sampling distribution of average life of Televisions by taking samples of size 2 and show that sample mean is an unbiased estimator of population mean life. Also show that S'^2 is not an unbiased estimator of population variance whereas S^2 is an unbiased estimator of population variance where,

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

Solution: Here, population consists three televisions whose lives are 8, 6 and 10 years so we can find the population mean and variance as

$$\mu = \frac{8+6+10}{3} = 8$$

$$\sigma^2 = \frac{1}{3}\left[(8-8)^2 + (6-8)^2 + (10-8)^2\right] = \frac{8}{3} = 2.67$$

Here, we are given that

Population size = N = 3 and sample size = n = 2

Therefore, possible numbers of samples (with replacement) that can be drawn from this population are $N^n = 3^2 = 9$. For each of these 9 samples, we will calculate the values of $\bar{X}, S'^2$ and S^2 by the formulae given below:

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

and necessary calculations for these results are shown in Table 5.1 given below:

Table 5.1: Calculation for $\bar{X}$, S'^2 and S^2

Sample	Sample Observation	$\bar{X}$	$\sum_{i=1}^{2}\left(X_i - \bar{X}\right)^2$	S'^2	S^2
1	8, 8	8	0	0	0
2	8, 6	7	2	1	2
3	8,10	9	2	1	2
4	6, 8	7	2	1	2
5	6, 6	6	0	0	0
6	6, 10	8	8	4	8
7	10, 8	9	2	1	2
8	10, 6	8	8	4	8
9	10, 10	10	0	0	0
Total		72		12	24

We calculate $\bar{X}, S'^2$ and S^2 as

$$\bar{X}_1 = \frac{1}{2}(8+8) = 8,\ \bar{X}_2 = \frac{1}{2}(8+6) = 7, \ldots, \bar{X}_9 = \frac{1}{2}(10+10) = 10$$

$$S_1'^2 = \frac{1}{2}\left[(8-8)^2 + (8-8)^2\right] = 0,\ S_2'^2 = \frac{1}{2}\left[(8-7)^2 + (6-7)^2\right] = 1, \ldots,$$

$$S_9'^2 = \frac{1}{2}\left[(10-10)^2 + (10-10)^2\right] = 0$$

$$S_1^2 = \frac{1}{2-1}\left[(8-8)^2 + (8-8)^2\right] = 0,\ S_2^2 = \frac{1}{2-1}\left[(8-7)^2 + (6-7)^2\right] = 2, \ldots,$$

$$S_9^2 = \frac{1}{2-1}\left[(10-10)^2 + (10-10)^2\right] = 0$$

Form the Table 5.1, we have

$$E\left(\bar{X}\right) = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_i = \frac{1}{9}\times 72 = 8 = \mu$$

Hence, sample mean is unbiased estimator of population mean.

Also

$$E\left(S'^2\right) = \frac{1}{k}\sum_{i=1}^{k}S_i'^2 = \frac{1}{9}\times 12 = 1.33 \neq \sigma^2$$

Therefore, S'^2 is not an unbiased estimator of σ^2 whereas,

$$E\left(S^2\right) = \frac{1}{k}\sum_{i=1}^{k}S_i^2 = \frac{1}{9}\times 24 = 2.67 = \sigma^2$$

S^2 is unbiased estimator of σ^2.

Remark 1:

1. Unbiased estimators may not be unique. For example, sample mean and sample median are unbiased estimators of population mean of normal population.
2. Unbiased estimators do not always exist for all the parameters. For example, for a Bernoulli distribution (θ), there is no unbiased estimator for θ^2. Similarly, for a Poisson distribution (λ), there exists no unbiased estimator for $1/\lambda$.
3. If an estimator is unbiased for all types of distribution, then it is called an absolutely unbiased estimator. For example, sample mean is an absolutely unbiased estimator of population mean, if population mean exists.
4. If T_n and T_n^* are two unbiased estimator of parameter θ then $aT_n + (1-a)\,T_n^*$ is also unbiased estimator of θ where, 'a' $(0 \le a \le 1)$ is any constant.

For the better understanding of the unbiasedness try some exercises.

E5) If $X_1, X_2, \ldots, X_n$ is a random sample taken from Poisson distribution whose probability mass function is given by

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

then show that the sample mean $\bar{X}$ is unbiased estimator of λ.

E6) If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population whose pdf is

$$f(x,\theta) = e^{-(x-\theta)}; \quad \theta \le x < \infty;\ -\infty < \theta < \infty$$

then show that the sample mean is an unbiased estimator of $(1 + \theta)$.

One weakness of unbiasedness is that it requires only the average value of the estimator equals to the true value of population parameter. It does not require those values of the estimator to be reasonably close to the population parameter. For this reason, we require some other properties of good estimator as consistency, efficiency and sufficiency which are described in subsequent sections.

5.5 CONSISTENCY

In previous section, we have learnt about the unbiasedness. An estimator T is said to be unbiased estimator of parameter, say, θ if the mean of sampling distribution of estimator T is equal to the true value of the parameter θ. This concept was defined for a fixed sample size. In this section, we will learn about consistency which is defined for increasing sample size.

If $X_1, X_2, ..., X_n$ is a random sample of size n taken from a population whose probability density (mass) function is f(x,θ) where, θ is the population parameter then consider a sequence of estimators, say, $T_1 = t_1(X_1)$, $T_2 = t_2(X_1, X_2)$, $T_3 = t_3(X_1, X_2, X_3)$,…, $T_n = t_n(X_1, X_2, ..., X_n)$. A sequence of estimators is said to be consistent for parameter θ if the deviation of the values of estimator from the parameter tends to zero as the sample size increases. That means values of estimators tend to get closer to the parameter θ as sample size increases.

In other words, a sequence $\{T_n\}$ of estimators is said to be consistent sequence of estimators of θ if T_n converges to θ in probability, that is

$$T_n \xrightarrow{p} \theta \text{ as } n \to \infty \quad \text{for every } \theta \in \Theta \qquad \dots (3)$$

or for every $\varepsilon > 0$

$$\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right] = 1 \qquad \dots (4)$$

or for every $\varepsilon > 0$ and $\eta > 0$ there exist $n \geq m$ such that

$$P\left[\left|T_n - \theta\right| < \varepsilon\right] > 1 - \eta \;; \qquad n \geq m \qquad \dots (5)$$

where, m is some very large value of n. Expressions (3), (4) and (5) are to mean the same thing.

Generally, to show that an estimator is consistent with the help of above definition is slightly difficult, therefore, we use sufficient conditions for consistency which are given below:

Sufficient conditions for consistency

If $\{T_n\}$is a sequence of estimators such that for all $\theta \in \Theta$

(i) $E(T_n) \to \theta$ as $n \to \infty$, that is, estimator T_n is either unbiased or asymptotically unbiased estimator of θ and

(ii) $\text{Var}(T_n) \to 0$ as $n \to \infty$, that is, variance of estimator T_n converges to 0 as $n \to \infty$.

Then estimator T_n is a consistent estimator of θ.

Now, we explain the procedure based on both the criteria (definition and sufficient condition) to show that a statistic is consistent or not for a parameter with the help of some examples:

Example 5: Prove that sample mean is always a consistent estimator of the population mean provided that the population has a finite variance.

Solution: Let $X_1, X_2, ..., X_n$ be a random sample taken from a population having mean μ and finite variance σ^2. By the definition of consistency, we have

$$\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right] = \lim_{n\to\infty} P\left[\left|\bar{X} - \mu\right| < \varepsilon\right] \quad \left[\text{Here}, T_n = \bar{X}\right]$$

$$= \lim_{n\to\infty} P\left[\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < \frac{\varepsilon\sqrt{n}}{\sigma}\right]$$

By central limit theorem (described in Unit 1 of this course), we know that the variate $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal variate for large sample size n. Therefore,

$$\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right] = \lim_{n\to\infty} P\left[\left|Z\right| < \frac{\varepsilon\sqrt{n}}{\sigma}\right]$$

$$= \lim_{n\to\infty} P\left[\frac{-\varepsilon\sqrt{n}}{\sigma} < Z < \frac{\varepsilon\sqrt{n}}{\sigma}\right] \quad \left[\because \left|X\right| < a \Rightarrow -a < X < a\right]$$

$$= \lim_{n\to\infty} \int_{\frac{-\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} f(z)\,dz \qquad \left[\because\ P\left[a < U < b\right] = \int_a^b f(u)\,du\right]$$

$$= \lim_{n\to\infty} \int_{-\frac{\varepsilon\sqrt{n}}{\sigma}}^{\frac{\varepsilon\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz \qquad \left[\because f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\right]$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$

Since $\dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is the pdf of a standard normal variate Z therefore, the integration of this in whole range $-\infty$ to ∞ is unity.

Thus,

$$\lim_{n\to\infty} P\left[\left|T_n - \theta\right| < \varepsilon\right] = \lim_{n\to\infty} P\left[\left|\bar{X} - \mu\right| < \varepsilon\right] = 1 \text{ as } n \to \infty$$

Hence, sample mean is a consistent estimator of population mean.

Note 2: This example can also be proved with the help of sufficient conditions for consistency as shown in next example.

Example 6: If $X_1, X_2, ..., X_n$ is a random sample taken from Poisson distribution(λ), then show that sample mean $\left(\bar{X}\right)$ is consistent estimator of λ.

Solution: We know that the mean and variance of Poisson distribution (λ) are

$$E(X) = \lambda \text{ and } Var(X) = \lambda$$

Since $X_1, X_2, ..., X_n$ are independent and come from same Poisson distribution, therefore,

$$E(X_i) = E(X) = \lambda \text{ and } Var(X_i) = Var(X) = \lambda \quad \text{for all } i = 1, 2, ..., n$$

Now consider,

$$E\left(\bar{X}\right) = E\left[\frac{1}{n}\left(X_1 + X_2 + ... + X_n\right)\right] \quad \left[\text{By defination of sample mean}\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \begin{bmatrix} \because E(aX + bY) \\ = aE(X) + bE(Y) \end{bmatrix}$$

$$= \frac{1}{n}\left(\underbrace{\lambda + \lambda + ... + \lambda}_{n-\text{times}}\right) = \frac{1}{n}(n\lambda) = \lambda$$

Thus, sample mean $\left(\bar{X}\right)$ is unbiased estimator of λ.

Now consider,

$$Var\left(\bar{X}\right) = Var\left[\frac{1}{n}\left(X_1 + X_2 + ... + X_n\right)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right] \quad \begin{bmatrix} \text{If X and Y are two} \\ \text{independent} \\ \text{random variable then} \\ Var(aX + bY) \\ = a^2Var(X) + b^2Var(Y) \end{bmatrix}$$

$$= \frac{1}{n^2}\left(\underbrace{\lambda + \lambda + ... + \lambda}_{n-\text{times}}\right)$$

$$= \frac{1}{n^2}(n\lambda)$$

$$Var\left(\bar{X}\right) = \frac{\lambda}{n} \to 0 \text{ as } n \to \infty$$

Hence, by sufficient condition of consistency, it follows that sample mean $(\bar{X})$ is consistent estimator of λ.

Remark 2:

1. Consistent estimators may not be unique. For example, sample mean and sample median are consistent estimators of population mean of normal population.
2. An unbiased estimator may or may not be consistent.
3. A consistent estimator may or may not be unbiased.

Continuous function is described in Sections 5.7 and 5.8 of Unit 5 of MST- 001.

4. If T_n is a consistent estimator of θ and f is a continuous function of θ then $f(T_n)$ is consistent estimator of f (θ). This property is known as invariance property. For example, if $\bar{X}$ is a consistent estimator of population mean θ then $e^{\bar{X}}$ also consistent estimator of e^{θ} because e^{θ} is a continuous function of θ.

It is now time for you to try the following exercises to make sure that you have understood consistency.

E7) If $X_1, X_2, ..., X_n$ is a random sample of size n taken from pdf

$$f(x,\theta) = \begin{cases} 1; & \theta \le x \le \theta + 1 \\ 0; & \text{elsewere} \end{cases}$$

then show that the sample mean is an unbiased as well as consistent estimator of $\left(\theta + \frac{1}{2}\right)$.

E8) If $X_1, X_2, ..., X_n$ are n observations taken from geometric distribution with parameter θ, then show that $\bar{X}$ is consistent estimator of 1/θ. Also find consistent estimator of $e^{1/\theta}$.

5.6 EFFICIENCY

In some situations, we see that there are more than one estimators of an parameter which are unbiased as well as consistent. For example, sample mean and sample median both are unbiased and consistent for the parameter μ when sampling is done from normal population with mean μ and known variance σ^2. In such situations, there arises a necessity of some other criterion which will help us to choose 'best estimator' among them. A criterion which is based on the concept of variance of the sampling distribution of the estimator is termed as efficiency.

If T_1 and T_2 are two estimators of an parameter θ. Then T_1 is said to be more efficient than T_2 for all sample sizes if

$$\text{Var}(T_1) < \text{Var}(T_2) \quad \text{for all n}$$

Let us do some examples about efficiency:

Example 7: Show that sample mean is more efficient estimator than sample median for estimating mean of normal population.

Solution: Let $X_1, X_2, ..., X_n$ be a random sample taken from normal population with mean μ and variance σ^2. Also let $\bar{X}$ and $\tilde{X}$ be the sample mean and sample median respectively. We have seen in Unit 2 that the sampling distribution of the mean from a normal population follows normal distribution with means μ and variance σ^2/n. Similarly, it can be shown that the sampling distribution of the median from a normal population also follows normal distribution with mean μ and variance $\frac{\pi}{2}\frac{\sigma^2}{n}$. Therefore,

$$\operatorname{Var}\left(\bar{X}\right)=\frac{\sigma^2}{n}$$

$$\operatorname{Var}\left(\tilde{X}\right)=\frac{\pi\sigma^2}{2n}$$

But $\frac{\pi\sigma^2}{2n}>\frac{\sigma^2}{n}\left[\because \frac{\pi}{2} \text{ and } \frac{\sigma^2}{n}>1\right]$ therefore, $\operatorname{Var}\left(\bar{X}\right)<\operatorname{Var}\left(\tilde{X}\right)$. Thus, we conclude that sample mean is more efficient estimator than sample median.

Example 8: If X_1, X_2, X_3, X_4 and X_5 is a random sample of size 5 taken from a population with mean μ and variance σ^2. The following two estimators are suggested to estimate μ

$$T_1=\frac{X_1+X_2+X_3+X_4+X_5}{5},\ T_2=\frac{X_1+2X_2+3X_3+4X_4+5X_5}{15}$$

Are both estimators unbiased? Which one is more efficient?

Solution: Since $X_1, X_2, \ldots, X_5$ are independent and taken from same population with mean μ and variance σ^2 therefore,

$$E\left(X_i\right)=\mu \text{ and } \operatorname{Var}\left(X_i\right)=\sigma^2 \qquad \text{for all } i=1,2,\ldots,5$$

Consider,

$$E\left(T_1\right)=E\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right]$$

$$=\frac{1}{5}\left[E\left(X_1\right)+E\left(X_2\right)+E\left(X_3\right)+E\left(X_4\right)+E\left(X_5\right)\right]$$

$$=\frac{1}{5}\left[\mu+\mu+\mu+\mu+\mu\right]$$

$$E\left(T_1\right)=\mu$$

Similarly,

$$E\left(T_2\right)=E\left[\frac{X_1+2X_2+3X_3+4X_4+5X_5}{15}\right]$$

$$=\frac{1}{15}\left[E\left(X_1\right)+2E\left(X_2\right)+3E\left(X_3\right)+4E\left(X_4\right)+5E\left(X_5\right)\right]$$

$$=\frac{1}{15}\left[\mu+2\mu+3\mu+4\mu+5\mu\right]$$

$$=\frac{1}{15}\left(15\mu\right)$$

$$E\left(T_2\right)=\mu$$

Hence, both the estimators T_1 and T_2 are unbiased estimators of μ.

Now for efficiency, we consider

$$\operatorname{Var}\left(T_1\right)=\operatorname{Var}\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right]$$

$$= \frac{1}{25}\left[\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) + \text{Var}(X_4) + \text{Var}(X_5)\right]$$

$$= \frac{1}{25}\left[\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2\right]$$

$$= \frac{1}{25}\left(5\sigma^2\right)$$

$$\text{Var}(T_1) = \frac{1}{5}\sigma^2$$

Similarly,

$$\text{Var}(T_2) = \text{Var}\left[\frac{X_1 + 2X_2 + 3X_3 + 4X_4 + 5X_5}{15}\right]$$

$$= \frac{1}{225}\begin{bmatrix}\text{Var}(X_1) + 4\text{Var}(X_2) + 9\text{Var}(X_3) \\ +16\text{Var}(X_4) + 25\text{Var}(X_5)\end{bmatrix}$$

$$= \frac{1}{225}\left[\sigma^2 + 4\sigma^2 + 9\sigma^2 + 16\sigma^2 + 25\sigma^2\right] = \frac{55\sigma^2}{225}$$

$$\text{Var}(T_2) = \frac{11\sigma^2}{45}$$

Since, $\text{Var}(T_1) < \text{Var}(T_2)$, therefore, we conclude that estimator T_1 is more efficient than T_2.

5.6.1 Most Efficient Estimator

In a class of estimators of a parameter, if there exists one estimator whose variance is minimum (least) among the class, then it is called most efficient estimator of that parameter. For example, suppose T_1, T_2 and T_3 are three estimators of parameter θ having variance $1/n$, $1/(n+1)$ and $5/n$ respectively. Since variance of estimator T_2 is minimum, therefore, estimator T_2 is most efficient estimator in that class.

The efficiency of an estimator measured with respect to the most efficient estimator is called **"Absolute Efficiency"**. If T^* is the most efficient estimator having variance $\text{Var}(T^*)$ and T is any other estimator having variance Var(T), then efficiency of T is defined as

$$e = \frac{\text{Var}(T^*)}{\text{Var}(T)}$$

Since variance of most efficient estimator is minimum, therefore,

$$e = \frac{\text{Var}(T^*)}{\text{Var}(T)} < 1$$

5.6.2 Minimum Variance Unbiased Estimator

An estimator T of parameter θ is said to be minimum variance unbiased estimator (MVUE) of θ if and only if

(i) $E(T) = \theta$ for all $\theta \in \Theta$, that is, estimator T is unbiased estimator of θ and

(ii) $\text{Var}(T) \le \text{Var}(T')$ for all $\theta \in \Theta$, that is, variance of estimator T is less than or equal to variance of any other unbiased estimator T'.

The minimum variance unbiased estimator (MVUE) is the most efficient unbiased estimator of parameter θ in the sense that it has minimum variance in class of unbiased estimators. Some authors used uniformly minimum variance unbiased estimator (UMVUE) in place of minimum variance unbiased estimator (MVUE).

Now, you can try the following exercises.

E9) If $X_1, X_2, ..., X_n$ is a random sample taken from a population having mean μ and variance σ^2, then show that the statistic $T' = \frac{1}{n+1}\sum_{i=1}^{n} X_i$ is biased but more efficient than sample mean for estimating the population mean.

E10) Suppose $X_1, X_2, ..., X_n$ is a random sample taken from normal population with mean μ and variance σ^2. The following two estimators are suggested to estimate μ as

$$T_1 = \frac{X_1 + X_2 + X_3}{3} \quad \text{and} \quad T_2 = \frac{X_1 + X_2}{2} + X_3$$

Are both estimators unbiased? Which one of them is more efficient?

E11) Define most efficient estimator and minimum variance unbiased estimator.

5.7 SUFFICIENCY

In statistical inference, the aim of the investigator or statistician may be to make a decision about the value of the unknown parameter (θ). The information that guides the investigator in making a decision is supplied by the random sample $X_1, X_2, ..., X_n$. However, in most of the cases the observations would be to numerous and too complicated. Directly use of these observations is complicated or cumbersome, therefore, a simplification or condensation would be desirable. The technique of condensing or reducing the random sample $X_1, X_2, ..., X_n$ into a statistic such that it contains all the information about parameter θ that is contained in the sample is known as sufficiency. So prior to continuing our search of finding best estimator, we introduce the concept of sufficiency.

A sufficient statistic is a particular kind of statistic that condenses random sample $X_1, X_2, ..., X_n$ in a statistic $T = t(X_1, X_2, ..., X_n)$ in such a way that no information about parameter θ is lost. That means, it contains all the information about θ that is contained in the sample and if we know the value of sufficient statistic, then the sample values themselves are not needed and can nothing tell you more about θ. In other words, A statistic T is said to be sufficient statistic for estimating a parameter θ if it contains all the information about θ which are available in the sample. This property of an estimator is called sufficiency. In other words,

An estimator T is sufficient for parameter θ if and only if the conditional distribution of $X_1, X_2, ..., X_n$ given T = t is independent of θ.

Mathematically,

$$f\left(x_1, x_2, ..., x_n / T = t\right) = g\left(x_1, x_2, ..., x_n\right)$$

where, the function $g\left(x_1, x_2, ..., x_n\right)$ does not depend on the parameter θ.

Note 3: Generally, the above definition is used to show that a particular statistic is not a sufficient statistic because it may be very tedious task to obtain the conditional distribution. Hence, we use factorization theorem which facilitates us to find sufficient statistic without any difficulty which is given below.

Theorem (Factorization Theorem): Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from the probability density (mass) function f(x, θ). A statistic or estimator T is said to be a sufficient for parameter θ if and only if the joint density (mass) function of $X_1, X_2, ..., X_n$ can be factored as

$$f\left(x_1, x_2, ..., x_n, \theta\right) = g\left[t(x), \theta\right].h\left(x_1, x_2, ..., x_n\right)$$

where, the function $g\left[t(x), \theta\right]$ is a non-negative function of parameter θ and observed sample values $(x_1, x_2, ..., x_n)$ only through the function t(x) and the function $h\left(x_1, x_2, ..., x_n\right)$ is non-negative function of $(x_1, x_2, ..., x_n)$ and does not involve the parameter θ.

For applying factorization theorem, we try to factor the joint density (mass) function as the product of two functions, one of which is function of parameter(s) and another is independent of parameter(s).

The proof of this theorem is beyond the scope of this course.

Note 4: The factorization theorem should not be used to show that a given statistic or estimator T is not sufficient.

Now, do some examples to show that a statistic is sufficient estimator for a parameter by using the factorization theorem:

Example 9: Show that the sample mean is sufficient for the parameter λ of the Poisson distribution.

Solution: We know that the probability mass function of Poisson distribution with parameter λ is

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

Let $X_1, X_2, ..., X_n$ be a random sample taken from Poisson distribution with parameter λ. Then joint mass function of $X_1, X_2, ..., X_n$ can be obtained as

$$f\left(x_1, x_2, ..., x_n, \lambda\right) = P\left[X_1 = x_1\right].P\left[X_2 = x_2\right] ... P\left[X_n = x_n\right]$$

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\cdot\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

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

$$= \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!} \qquad \left[\prod_{i=1}^{n}x_i! - \text{repersents the product of } x_i!\right]$$

The joint mass function can be factored as

$$f(x_1, x_2, \ldots, x_n, \lambda) = \left(e^{-n\lambda}\lambda^{n\bar{x}}\right)\times\left(\frac{1}{\prod_{i=1}^{n}x_i!}\right) \qquad \left[\because\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i\right]$$

$$= g[t(x), \lambda].h(x_1, x_2, \ldots, x_n)$$

where, $g[t(x), \lambda] = e^{-n\lambda}\lambda^{n\bar{x}}$ is a function of parameter λ and the observed sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \bar{x}$ and $h(x_1, x_2, \ldots, x_n) = \dfrac{1}{\prod_{i=1}^{n}x_i!}$ is a function of sample values $x_1, x_2, \ldots, x_n$ and is independent of parameter λ.

Hence, by factorization theorem of sufficiency, $\bar{X}$ is sufficient statistic for λ.

Note 4: Since throughout the course we are using capital letter for statistic or estimator therefore in the last line of above example we use $\bar{X}$ in place of $\bar{x}$. Thus, in all the examples and exercises relating to sufficient statistic, we are using similar approach.

Example 10: A random sample $X_1, X_2, \ldots, X_n$ of size n is taken from gamma distribution whose pdf is given by

$$f(x, a, b) = \frac{a^b}{\overline{|b}}e^{-ax}x^{b-1}$$

Obtain sufficient statistic for

(i) 'b' when 'a' is known

(ii) 'a' when 'b' is known

(iii) 'a' and 'b' both are unknown.

Solution: The joint density function can be obtained as

$$f(x_1, x_2, \ldots, x_n, a, b) = f(x_1, a, b)f(x_2, a, b)\ldots f(x_n, a, b)$$

$$= \frac{a^b}{\overline{|b}}e^{-ax_1}x_1^{b-1}.\frac{a^b}{\overline{|b}}e^{-ax_2}x_2^{b-1}\ldots\frac{a^b}{\overline{|b}}e^{-ax_n}x_n^{b-1}$$

$$= \frac{a^{nb}}{\left(\overline{|b}\right)^n}e^{-a\sum_{i=1}^{n}x_i}\left(\prod_{i=1}^{n}x_i\right)^{b-1}$$

Since 'a' is known so 'a' is treated as a constant.

Case I: When 'a' is known then we find sufficient statistic for 'b'

The joint density function can be factored as

$$f\left(x_1, x_2, \ldots, x_n, b\right) = \left[\frac{a^{nb}}{\left(\overline{|b}\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}\right].e^{-a\sum_{i=1}^{n} x_i}$$

$$= g\left[t(x), b\right].h\left(x_1, x_2, \ldots x_n\right)$$

where, $g\left[t(x), b\right] = \frac{a^{nb}}{\left(\overline{|b}\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}$ is a function of parameter 'b' and sample

values $x_1, x_2, \ldots, x_n$ only through $t(x) = \prod_{i=1}^{n} x_i$ and $h\left(x_1, x_2, \ldots, x_n\right) = e^{-a\sum_{i=1}^{n} x_i}$ is a function of sample values $x_1, x_2, \ldots, x_n$ and is independent of parameter 'b'.

Hence, by factorization theorem of sufficiency, $\prod_{i=1}^{n} X_i$ is sufficient statistic for parameter 'b'

Case II: When 'b' is known then we find sufficient statistic for 'a'

The joint density function can be factored as

Since 'b' is known so 'b' is treated as a constant.

$$f\left(x_1, x_2, \ldots, x_n, a\right) = \left[a^{nb} e^{-a\sum_{i=1}^{n} x_i}\right].\frac{1}{\left(\overline{|b}\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}$$

$$= g\left[t(x), a\right].h\left(x_1, x_2, \ldots x_n\right)$$

where, $g\left[t(x), a\right] = a^{nb} e^{-a\sum_{i=1}^{n} x_i}$ is a function of parameter 'a' and sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \sum_{i=1}^{n} x_i$ and $h\left(x_1, x_2, \ldots, x_n\right) = \frac{1}{\left(\overline{|b}\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}$ is a function of sample values $x_1, x_2, \ldots, x_n$ and is independent of parameter 'a'.

Hence by factorization theorem of sufficiency, $\sum_{i=1}^{n} X_i$ is sufficient statistic for 'a'.

Case III: When 'a' and 'b' are unknown then we find jointly sufficient statistics for 'a' and 'b'

The joint density function can be factored as

$$f\left(x_1, x_2, \ldots, x_n, a, b\right) = \left[a^{nb} e^{-a\sum_{i=1}^{n} x_i} \frac{1}{\left(\overline{|b}\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}\right].1$$

$$= g\left[t_1(x), t_2(x), a, b\right].h\left(x_1, x_2, \ldots, x_n\right)$$

where, $g\left[t_1(x), t_2(x), a, b\right] = a^{nb} e^{-a\sum_{i=1}^{n} x_i} \dfrac{1}{\left(\lceil b\right)^n}\left(\prod_{i=1}^{n} x_i\right)^{b-1}$ is a function of parameters 'a'& 'b' and sample values $x_1, x_2, ..., x_n$ only through $t_1(x) = \sum_{i=1}^{n} x_i$ and $t_2(x) = \prod_{i=1}^{n} x_i$ whereas, $h\left(x_1, x_2, ... x_n\right) = 1$ and independent of parameters 'a' and 'b'.

Hence, by factorization theorem, $\sum_{i=1}^{n} X_i$ and $\prod_{i=1}^{n} X_i$ are jointly sufficient for parameters 'a'& 'b'.

Example 11: If $X_1, X_2, ..., X_n$ is a random sample taken from uniform distribution U(α, β), find the sufficient statistics for α and β.

Solution: The probability density function of U(α, β) is given by

$$f(x, \alpha, \beta) = \frac{1}{\beta - \alpha}; \quad \alpha \le x \le \beta$$

The joint density function can be obtained as

$$f\left(x_1, x_2, ..., x_n, \alpha, \beta\right) = f\left(x_1, \alpha, \beta\right).f\left(x_2, \alpha, \beta\right) ... f\left(x_n, \alpha, \beta\right)$$

$$= \frac{1}{\beta-\alpha}.\frac{1}{\beta-\alpha}...\frac{1}{\beta-\alpha}$$

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

cs of
e
e the
aced
er of
e are
$X_{(n)}$

Since the range of variable depends upon the parameters so we consider ordered statistics $X_{(1)}, X_{(2)}, ..., X_{(n)}$. Therefore, the joint density function can be factored as

$$f\left(x_1, x_2, ..., x_n, \alpha, \beta\right) = \frac{1}{(\beta-\alpha)^n}; \qquad \alpha \le x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \le \beta$$

$$= \left[\frac{1}{(\beta-\alpha)^n} I_1\left(x_{(1)}, \alpha\right) I_2\left(x_{(n)}, \beta\right)\right].1$$

where, $x_{(1)}$ and $x_{(n)}$ are the minimum and maximum sample observations respectively and

$$I_1\left(x_{(1)}, \alpha\right) = \begin{cases} 1; & \text{if } x_{(1)} \ge \alpha \\ 0; & \text{otherwise} \end{cases}$$

$$I_2\left(x_{(n)}, \beta\right) = \begin{cases} 1; & \text{if } x_{(n)} \le \beta \\ 0; & \text{otherwise} \end{cases}$$

Therefore,

$$f\left(x_1, x_2, ..., x_n, \alpha, \beta\right) = g\left[t_1(x), t_2(x), \alpha, \beta\right].h\left(x_1, x_2, ..., x_n\right)$$

where, $g[t_1(x), t_2(x), \alpha, \beta] = \left[\frac{1}{(\beta-\alpha)^n} I_1\left(x_{(1)}, \alpha\right) I_2\left(x_{(n)}, \beta\right)\right]$ is a function of parameters (α, β) and sample values $x_1, x_2, ..., x_n$ only through $t_1(x) = x_{(1)}$ and $t_2(x) = x_{(n)}$ whereas, $h\left(x_1, x_2, ... x_n\right) = 1$ and independent of parameters 'α' and 'β'.

Hence, by factorization theorem of sufficiency, $X_{(1)}$ and $X_{(n)}$ are jointly sufficient for α and β.

Remark 3:

1. A sufficient estimator is always a consistent estimator.
2. A sufficient estimator may be unbiased.
3. A sufficient estimator is the most efficient estimator if an efficient estimator exists.
4. The random sample $X_1, X_2, ..., X_n$ and order statistics $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are always sufficient estimators because both are contain all the information about the parameter(s) of the population.
5. If T is a sufficient statistic for the parameter θ and ϕ(T) is a one to one function of T then ϕ(T) is also sufficient for θ. For example, if $T = \sum X_i$ is sufficient statistic for parameter θ then $\bar{X} = \frac{1}{n}\sum X_i = \frac{T}{n}$ is also sufficient for θ because $\bar{X} = \frac{T}{n}$ is a one to one function of T.

Now, you will understand more clearly about the sufficiency, when you try the following exercises.

E12) If $X_1, X_2, ..., X_n$ is a random sample taken from exp (θ) then find sufficient statistic for θ.

E13) If $X_1, X_2, ..., X_n$ is a random sample taken from normal population N(μ, σ^2), then obtain sufficient statistic for μ and σ^2 or both according as other parameter is known or unknown.

E14) If $X_1, X_2, ..., X_n$ is a random sample from uniform population over the interval [0, θ]. Find sufficient estimator of θ.

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

5.8 SUMMARY

In this unit, we have covered the following points:

1. The parameter space and joint probability density (mass) function.
2. The basic characteristics of an estimator.
3. Unbiasedness of an estimator.
4. Consistency of an estimator.
5. Efficiency of an estimator.

6. The most efficient estimator.
7. Minimum variance unbiased estimator.
8. The sufficiency of an estimator.

5.9 SOLUTIONS / ANSWERS

E1) We know that the pmf of Poisson distribution is

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

and mean and variance of this distribution are

Mean = Variance = λ

In our case, $\lambda = 5$, therefore, pmf of Poisson distribution is

$$P\left[X=x\right]=\frac{e^{-5}5^{x}}{x!};\quad x=0,1,2,\ldots$$

Also mean and variance of this distribution are

Mean = Variance = λ = 5

E2) Since parameter θ represents the average marks of IGNOU's students in a paper of 50 marks, therefore, a student can take minimum 0 marks and maximum 50 marks. Thus, the parameter space of θ is $\Theta=\left\{\theta : 0\le\theta\le 50\right\}$.

E3) The probability mass function of Poisson distribution is

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

The joint probability mass function of $X_1, X_2, \ldots, X_n$ can be obtained as

$$f\left(x_1,x_2,\ldots,x_n,\lambda\right)=P\left[X_1=x_1\right].P\left[X_2=x_2\right]\ldots P\left[X_n=x_n\right]$$

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

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

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

E4) Refer Section 5.3.

E5) We know that the mean of Poisson distribution with parameter λ is λ i.e.

$$E(X)=\lambda$$

Since $X_1, X_2, ..., X_n$ are independent and come from same Poisson distribution, therefore,

$$E(X_i) = E(X) = \lambda \quad \text{for all } i = 1, 2, ..., n$$

Now consider,

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right] \qquad \begin{bmatrix}\text{By defination of}\\ \text{sample mean}\end{bmatrix}$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \begin{bmatrix}\because E(aX + bY)\\ = aE(X) + bE(Y)\end{bmatrix}$$

$$= \frac{1}{n}\left(\underbrace{\lambda + \lambda + ... + \lambda}_{n-\text{times}}\right) = \frac{1}{n}(n\lambda) = \lambda$$

Hence, sample mean $(\bar{X})$ is unbiased estimator of parameter λ.

E6) Here, we have to show that

$$E(\bar{X}) = 1 + \theta$$

We are given that

$$f(x, \theta) = e^{-(x-\theta)}; \quad \theta \le x < \infty, \ -\infty < \theta < \infty$$

Since we do not know the mean of this distribution therefore first of all we find the mean of this distribution. So we consider,

$$E(X) = \int_{\theta}^{\infty} x\, f(x, \theta)\, dx$$

$$= \int_{\theta}^{\infty} x e^{-(x-\theta)} dx$$

Putting $x - \theta = y \Rightarrow dx = dy$. Also when $x = \theta \Rightarrow y = 0$ & when $x \to \infty \Rightarrow y \to \infty$. Therefore,

$$E(X) = \int_{0}^{\infty} (y + \theta) e^{-y} dy$$

$$= \int_{0}^{\infty} y e^{-y} dy + \theta \int_{0}^{\infty} e^{-y} dy$$

$$= \int_{0}^{\infty} y^{2-1} e^{-y} dy + \theta \int_{0}^{\infty} y^{1-1} e^{-y} dy$$

$$= \overline{|2} + \theta \times \overline{|1} = 1 + \theta \qquad \begin{bmatrix}\because \int_{0}^{\infty} x^{n-1} e^{-x} dx = \overline{|n}\\ \text{and } \overline{|2} = \overline{|1} = 1\end{bmatrix}$$

Since $X_1, X_2, ..., X_n$ are independent and come from same population, so

$$E(X_i) = E(X) = 1 + \theta \quad \text{for all } i = 1, 2, ..., n$$

Now consider,

$$E(\bar{X}) = E\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] \quad \left[\begin{array}{l}\text{By defination of} \\ \text{sample mean}\end{array}\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \left[\begin{array}{l}\because E(aX + bY) \\ = aE(X) + bE(Y)\end{array}\right]$$

$$= \frac{1}{n}\left[\underbrace{(1+\theta) + (1+\theta) + ... + (1+\theta)}_{n-\text{times}}\right]$$

$$= \frac{1}{n}\left[n(1+\theta)\right]$$

$$= 1 + \theta$$

Thus, sample mean is an unbiased estimator of (1+θ).

E7) We have

$$f(x, \theta) = 1; \quad \theta \le x \le \theta + 1$$

This is the pdf of uniform distribution U[θ, θ+1] and we know that for U[a, b]

$$E(X) = \frac{a+b}{2} \text{ and } Var(X) = \frac{(b-a)^2}{12}$$

In our case, a = θ and b = θ + 1, therefore

$$E(X) = \frac{\theta + \theta + 1}{2} = \theta + \frac{1}{2} \text{ and } Var(X) = \frac{(\theta + 1 - \theta)^2}{12} = \frac{1}{12}$$

Since $X_1, X_2, ..., X_n$ are independent and come from same population, therefore,

$$E(X_i) = E(X) = \theta + \frac{1}{2} \text{ and } Var(X_i) = Var(X) = \frac{1}{12} \quad \forall\, i = 1, 2, ..., n$$

To show that $\bar{X}$ is unbiased estimator of $\left(\theta + \frac{1}{2}\right)$, we consider

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right] \quad \left[\begin{array}{l}\text{By defination of} \\ \text{sample mean}\end{array}\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + ... + E(X_n)\right] \quad \left[\begin{array}{l}\because E(aX + bY) \\ = aE(X) + bE(Y)\end{array}\right]$$

$$= \frac{1}{n}\left[\underbrace{\left(\theta + \frac{1}{2}\right) + \left(\theta + \frac{1}{2}\right) + ... + \left(\theta + \frac{1}{2}\right)}_{n-\text{times}}\right]$$

$$= \frac{1}{n}\left[n\left(\theta + \frac{1}{2}\right)\right]$$

$$= \theta + \frac{1}{2}$$

Therefore, $\bar{X}$ is unbiased estimator of $\left(\theta + \frac{1}{2}\right)$.

For consistency, we have to show that

$$E\left(\bar{X}\right) \to \theta + \frac{1}{2} \text{ and } Var\left(\bar{X}\right) \to 0 \text{ as } n \to \infty$$

If X and Y are two independent random variables and a & b are two constants then $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$

Now consider,

$$Var\left(\bar{X}\right) = Var\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right]$$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right]$$

$$= \frac{1}{n^2}\left(\underbrace{\frac{1}{12} + \frac{1}{12} + ... + \frac{1}{12}}_{n-times}\right)$$

$$= \frac{1}{n^2}\left(\frac{n}{12}\right)$$

$$\text{Now, } Var\left(\bar{X}\right) = \frac{1}{12n} \to 0 \text{ as } n \to \infty$$

$$\text{Thus, } E\left(\bar{X}\right) = \theta + \frac{1}{2} \text{ and } Var(\bar{X}) \to 0 \text{ as } n \to \infty$$

Hence, sample mean $\bar{X}$ is also consistent estimator of $\left(\theta + \frac{1}{2}\right)$.

E8) We know that the mean and variance of geometric distribution(θ) are given by

$$E(X) = \frac{1}{\theta} \text{ and } Var\left(X\right) = \frac{1-\theta}{\theta^2}$$

Since $X_1, X_2, ..., X_n$ are independent and come from same geometric distribution, therefore,

$$E(X_i) = E(X) \text{ and } Var\left(X_i\right) = Var\left(X\right) \qquad \text{for all } i = 1, 2, ..., n$$

First, we show sample mean $\bar{X}$ is consistent estimator of $1/\theta$.

Therefore, we consider,

$$E\left(\bar{X}\right) = E\left[\frac{1}{n}(X_1 + X_2 + ... + X_n)\right] \qquad \left[\begin{array}{l}\text{By defination of} \\ \text{sample mean}\end{array}\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right] \quad \left[\begin{array}{l}\because E(aX + bY) \\ = aE(X) + bE(Y)\end{array}\right]$$

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

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

Now consider,

$$\text{Var}\left(\bar{X}\right) = \text{Var}\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right]$$

$$= \frac{1}{n^2}\left[\text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n)\right]$$

$$= \frac{1}{n^2}\left[\underbrace{\left(\frac{1-\theta}{\theta^2}\right) + \left(\frac{1-\theta}{\theta^2}\right) + \ldots + \left(\frac{1-\theta}{\theta^2}\right)}_{n-\text{times}}\right]$$

$$= \frac{1}{n^2}\left[n\left(\frac{1-\theta}{\theta^2}\right)\right] = \frac{1}{n}\left(\frac{1-\theta}{\theta^2}\right)$$

$$\text{Var}\left(\bar{X}\right) = \frac{1}{n}\left(\frac{1-\theta}{\theta^2}\right) \to 0 \text{ as } n \to \infty$$

Since $E\left(\bar{X}\right) = \frac{1}{\theta}$ and $\text{Var}(\bar{X}) \to 0$ as $n \to \infty$

Hence, sample mean $\bar{X}$ is consistent estimator of $1/\theta$.

Since $e^{1/\theta}$ is continuous function of $1/\theta$ therefore, by invariance property of consistency $e^{\bar{X}}$ is consistent estimator of $e^{1/\theta}$.

E9) Since $X_1, X_2, \ldots, X_n$ is a random sample taken from a population having mean μ and variance σ^2.

Therefore,

$$E(X_i) = \mu \text{ and } \text{Var}(X_i) = \sigma^2 \qquad \text{for all } i = 1, 2, \ldots, n$$

Consider,

$$E(T') = E\left(\frac{1}{n+1}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n+1} E\left(X_1 + X_2 + \ldots + X_n\right)$$

$$= \frac{1}{n+1}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]\left[\begin{array}{l}\because E(aX + bY) \\ = aE(X) + bE(Y)\end{array}\right]$$

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

$$= \frac{1}{n+1}(n\mu) \neq \mu$$

Therefore, T' is biased estimator of population mean μ.

For efficiency, we find variances of estimator T' and sample mean $\bar{X}$ as

$$\text{Var}(T') = \text{Var}\left[\frac{1}{n+1}(X_1 + X_2 + \ldots + X_n)\right]$$

If X and Y are two independent random variables and a & b are two constants then
$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$

$$= \frac{1}{(n+1)^2}\left[\text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n)\right]$$

$$= \frac{1}{(n+1)^2}\left(\underbrace{\sigma^2 + \sigma^2 + \ldots + \sigma^2}_{n-\text{times}}\right)$$

$$= \frac{1}{(n+1)^2}\left(n\sigma^2\right) = \frac{n\sigma^2}{(n+1)^2}$$

Now consider,

$$\text{Var}(\bar{X}) = \text{Var}\left[\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right]$$

$$= \frac{1}{n^2}\left[\text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_n)\right]$$

$$= \frac{1}{n^2}\left(\underbrace{\sigma^2 + \sigma^2 + \ldots + \sigma^2}_{n-\text{times}}\right)$$

$$= \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}$$

Since, $\text{Var}(T') < \text{Var}(\bar{X})$, therefore, T' is more efficient estimator than sample mean.

E10) Since $X_1, X_2, \ldots, X_n$ are independent taken from normal population with mean μ and variance σ^2 therefore,

$$E(X_i) = \mu \text{ and } \text{Var}(X_i) = \sigma^2 \qquad \text{for all } i = 1, 2, 3$$

To check estimators T_1 and T_2 are unbiased, we find expectations of T_1 and T_2 as

If X and Y are two independent random variables and a & b are two constants then
$E(aX + bY) = aE(X) + bE(Y)$ and
$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$

$$E(T_1) = E\left(\frac{X_1 + X_2 + X_3}{3}\right)$$

$$= \frac{1}{3}\left[E(X_1) + E(X_2) + E(X_3)\right]$$

$$E(T_1) = \frac{1}{3}\left[\mu + \mu + \mu\right]$$

$$= \frac{1}{\mu}(3\mu) = \mu$$

Now consider,

$$E(T_2) = E\left(\frac{X_1 + X_2}{4} + \frac{X_3}{2}\right)$$

$$= \frac{1}{4}\left[E(X_1) + E(X_2)\right] + \frac{1}{2}E(X_3)$$

$$= \frac{1}{4}[\mu + \mu] + \frac{\mu}{2} = \frac{\mu}{2} + \frac{\mu}{2}$$

$$E(T_2) = \mu$$

Hence, T_1 and T_2 both are unbiased estimators of μ.

For efficiency, we find the variances of T_1 and T_2 as

$$\text{Var}(T_1) = \text{Var}\left(\frac{X_1 + X_2 + X_3}{3}\right)$$

$$= \frac{1}{9}\left[\text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3)\right]$$

$$= \frac{1}{9}\left[\sigma^2 + \sigma^2 + \sigma^2\right] = \frac{3\sigma^2}{9}$$

$$\text{Var}(T_1) = \frac{\sigma^2}{3}$$

Now consider,

$$\text{Var}(T_2) = \text{Var}\left(\frac{X_1 + X_2}{4} + \frac{X_3}{2}\right)$$

$$= \frac{1}{16}\left[\text{Var}(X_1) + \text{Var}(X_2)\right] + \frac{1}{4}\text{Var}(X_3)$$

$$= \frac{1}{16}\left[\sigma^2 + \sigma^2\right] + \frac{1}{4}\sigma^2$$

$$= \frac{\sigma^2}{8} + \frac{\sigma^2}{4} = \frac{\sigma^2 + 2\sigma^2}{8}$$

$$= \frac{3\sigma^2}{8}$$

Since $\text{Var}(T_1) < \text{Var}(T_2)$ therefore, T_1 is more efficient estimator of μ than T_2.

E11) Refer Sub-sections 5.6.1 and 5.6.2.

E12) Here, we take random sample from exp (θ) whose probability density function is given by

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

The joint density function of $X_1, X_2, ..., X_n$ can be obtained and can be factored as

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

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

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

$$= \left[\theta^n e^{-\theta\sum_{i=1}^{n} x_i}\right].1$$

$$L(\theta) = g\left[t(x), \theta\right].h(x_1, x_2, ..., x_n)$$

where, $g\left[t(x), \theta\right] = \theta^n e^{-\theta\sum_{i=1}^{n} x_i}$ is a function of parameter θ and sample values $x_1, x_2, ..., x_n$ only through $t(x) = \sum_{i=1}^{n} x_i$ and $h(x_1, x_2, ..., x_n) = 1$, is independent of θ. Hence by factorization theorem of sufficiency, $\sum_{i=1}^{n} X_i$ is sufficient estimator of θ.

E13) Here, we take random sample from N(μ, σ^2) whose probability density function is

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

The joint density function of $X_1, X_2, ..., X_n$ can be obtained as

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

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

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-\mu)]^2} \qquad = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i}$$

… (1)

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2} \quad [\text{add and subtract } \bar{x}]$$

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

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

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2-0\right]} \quad \left[\begin{array}{l}\sum_{i=1}^{n}(x_i-\bar{x})=0, \text{by}\\ \text{the property of mean}\end{array}\right]$$

$$f(x_1, x_2, \ldots x_n, \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} \qquad \ldots (2)$$

Case I: Sufficient statistic for μ when σ^2 is known

The joint density function given in equation (2) can be factored as

$$f(x_1, x_2, \ldots x_n, \mu) = \left[e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} \cdot \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\right] \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$= g[t(x), \mu].h(x_1, x_2, \ldots, x_n)$$

where, $g[t(x), \mu] = e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2} \cdot \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n$ is a function of parameter μ and sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \bar{x}$, whereas

$h(x_1, x_2, \ldots, x_n) = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ is independent of μ. Hence, $\bar{X}$ is sufficient estimator for μ when σ^2 is known.

Case II: Sufficient statistic for σ^2 when μ is known

The joint density function given in equation (1) can be factored as

$$f(x_1, x_2, \ldots x_n, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

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

$$= g[t(x), \mu].h(x_1, x_2, \ldots, x_n)$$

where, $g[t(x), \mu] = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$ is a function of parameter

σ^2 and sample values $x_1, x_2, \ldots, x_n$ only through $t(x) = \sum_{i=1}^{n}(x_i-\mu)^2$, whereas $h(x_1, x_2, \ldots, x_n) = 1$ is independent of σ^2. Hence by factorization theorem of sufficiency, $\sum_{i=1}^{n}(X_i-\mu)^2$ is sufficient estimator for σ^2 when μ is known.

Case III: When both μ and σ^2 are unknown

The joint density function given in equation (2) can be factored as

$$f(x_1, x_2, \ldots x_n, \mu, \sigma^2) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left[(n-1)s^2-n(\bar{x}-\mu)^2\right]}\right].1$$

where, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$

$f(x_1, x_2, ... x_n, \mu, \sigma^2) = g(t_1(x), t_2(x), \mu, \sigma^2).h(x_1, x_2, ..., x_n)$

where, $g(t_1(x), t_2(x), \mu, \sigma^2)$ is a function of (μ, σ^2) and sample values $x_1, x_2, ..., x_n$ only through $t_1(x) = \bar{x}, t_2(x) = s^2$, whereas $h(x_1, x_2, ..., x_n)$ is independent of (μ, σ^2). Hence by factorization theorem, $\bar{X}$ and S^2 are jointly sufficient for μ and σ^2.

Note 5: Here, it is remembered that $\bar{X}$ is not sufficient statistic for μ if σ^2 is unknown and S^2 is not sufficient for σ^2 if μ is unknown.

E14) The sample is taken from U [0,θ] whose probability density function is

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

The joint density function of $X_1, X_2, ..., X_n$ can be obtained as

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

$$= \frac{1}{\theta}.\frac{1}{\theta} ... \frac{1}{\theta} = \frac{1}{\theta^n}$$

Since range of variable depends upon the parameter θ, so we consider ordered statistics $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$.

Therefore, the joint density function can be factored as

$$f(x_1, x_2, ... x_n, \theta) = \frac{1}{\theta^n}; \qquad 0 \le x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \le \theta$$

$$= \left[\frac{1}{\theta^n} I(x_{(n)}, \theta)\right].1$$

where,

$$I(x_{(n)}, \theta) = \begin{cases} 1; & \text{if } x_{(n)} \le \theta \\ 0; & \text{otherwise} \end{cases}$$

Therefore,

$$f(x_1, x_2, ... x_n, \theta) = g[t(x), \theta].h(x_1, x_2, ..., x_n)$$

where, $g[t(x), \theta]$ is a function of θ and sample values only through $t(x) = x_{(n)}$, whereas $h(x_1, x_2, ..., x_n)$ is independent of θ.

Hence, by factorization theorem $X_{(n)}$ is sufficient statistic for θ.